A PUBLICATION OF THE AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Abnormal and Social Psychology

THE UNIVERSITY OF MICHIGAN
SEP 22 1961
UNDERGRADUATE LIBRARY

Founded by Morton Prince, 1906

CONTENTS

ARTICLES

"Credulous" and "Skeptical" Views of Hypnotic Phenomena: Experiments on Esthesia, Hallucination, and Delusion 189
J. P. Sutcliffe
Performance on Complex Reasoning Tasks as a Function of Anxiety 201
Charles Van Buskirk
Anti-Semitism, Judgmental Processes, and Displacement of Hostility 210
Leonard Berkowitz
The Effect of Amount of Possession of the Trait of Hostility on Accuracy of Perception of Hostility in Others 216
Bernard I. Murstein
Age Differences: A Developmental Study of Discriminability and Affective Response 221
Nathan Kogan, Judith W. Stephens, and Florence C. Shelton
Cognitive Risk and Environmental Change 231
Jerome S. Bruner and Henri Tajfel
Existential Psychotherapy—Fact, Hope, Fiction 242
Joseph Lyons
Generalization of Interpersonal Similarity 250
Ezra Stotland, Alvin Zander, and Thomas Natsoulas
Similarity to a Model and Self-Evaluation 257
Eugene Burnstein, Ezra Stotland, and Alvin Zander
Identification and Changes in Prejudice and in Authoritarianism 265
Ezra Stotland and Martin Patchen
On Group Composition 275
William C. Schutz
An Investigation of Unilateral Cerebral Pathology with Equated Verbal and Visual-Spatial Tasks 282
Rosemary Stark
Some Determinants of Self-Evaluation 288
Harold B. Gerard
The Social Psychological Nature of Stress: A Framework for Causal Inquiry 294
Bruce P. Dohrenwend

Continued on inside front cover

Volume 62 MARCH 1961 Number 2

CONTENTS—*Continued*

Individual Differences in Reaction to Failure: Mode of Coping with Anxiety and Interference Proneness ... 303
Alfred Lowe

Self-Concept, Mother Concept, and Food Aversions in Emotionally Disturbed and Normal Children ... 309
Anthony Davids and Marcia J. Lawton

Children's Attitudes toward Rational versus Inhibiting Parental Authority ... 315
Anatol Pikas

Paired-Associate Learning by Schizophrenic and Normal Subjects under Conditions of Personal and Impersonal Reward and Punishment ... 322
Rita L. Atkinson and Nancy M. Robinson

The Relative Efficacy of Various Types of Prior Belief-Defense in Producing Immunity against Persuasion ... 327
William J. McGuire and Demetrios Papageorgis

Affective Tone during Environmental Regularity and Change ... 338
Salvatore R. Maddi

The Effects of Frustration and Anxiety on Overt Aggression ... 346
Jack E. Hokanson

The Performance of Psychopathic Individuals under Conditions of Positive and Negative Partial Reinforcement ... 352
Donald H. Painting

Anxiety, Fear, and Social Affiliation ... 356
Irving Sarnoff and Philip G. Zimbardo

Microgenetic Analysis of Word Associations in Schizophrenic and Brain Damaged Patients ... 364
Norman A. Milgram

A Free Response Approach to the Study of Person Cognition ... 367
LeeRoy Beach and Michael Wertheimer

Address in American English ... 375
Roger Brown and Marguerite Ford

What Do Attitude Scales Measure: The Problem of Social Desirability ... 386
James Bentley Taylor

The Interpretation of Praise and Criticism as a Function of Their Social Context ... 391
Morton Deutsch

Quality and Acceptance of Problem Solutions by Members of Homogeneous and Heterogeneous Groups ... 401
L. Richard Hoffman and Norman R. F. Maier

Group Cohesiveness, Communication Level, and Conformity ... 408
Albert J. Lott and Bernice Eisman Lott

Social Deprivation and Rigidity in the Performance of Feebleminded Children ... 413
Edward Zigler

CASE REPORT

Motivational Structure of an Hospitalized Mental Defective ... 422
Anna M. Shotwell, John R. Hurley, and Raymond B. Cattell

CRITIQUE AND NOTES

GSR Conditioning, Anxiety, and Extraversion ... 427
Wesley C. Becker and Halsey H. Matteson

Relationship of Persuasive Interaction to Change of Opinion in Dyadic Groups when the Original Opinions of Participants are Expressed Privately and Publicly ... 431
Vladimir B. Cervin, Robert C. Joyner, J. Michael Spence, and Rudolph Heinzl

Personal Constructs and the Semantic Differential ... 433
Spiro B. Mitsos

Incongruency as a Predictor of Response to Humor ... 435
Donn Byrne, James Terrill, and Paul McReynolds

Perceived Sources of Social Power ... 439
Sidney Rosen, George Levinger, and Ronald Lippitt

Interpersonal Perception Scores and Their Components as Predictors of Personal Adjustment 442
Alan R. Bass and Fred E. Fiedler

The Generality of Accuracy of Interpersonal Perception ... 446
Victor B. Cline and James M. Richards, Jr.

The Effects of Prior Information on Susceptibility to an Emotional Appeal ... 450
Paul C. Lewan and Esra Stotland

Anti-Semitism and the Expression of Fantasy Aggression ... 454
Donald Weatherley

Some Comments on the EPPS Ego Disjunction Score ... 458
Edward F. Gocka and Vitali Rozynko

Effects of Aggressive Cartoons on Children's Aggressive Play ... 461
Paul Mussen and Eldred Rutherford

Relationships between Hope of Success and Fear of Failure, Anxiety, and Need for Achievement ... 465
E. Edward Reitman and Carl D. Williams

The Relationship between Anxiety and *Oral* Word Association Performance ... 468
Michael J. Goldstein

Relationships between Self-Acceptance and Sociometric Choices ... 472
Hayne W. Reese

303
309
315
322
327
338
346
352
356
364
367
375
386
391
401
408
413
422
427
431
433
435
439
442
446
450
454
458
461
465
468
472

Volume 62 March 1961 Number 2

Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *University of California, Berkeley*
Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

N. L. GAGE
University of Illinois

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

RICHARD S. LAZARUS
University of California, Berkeley

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California, Berkeley

ELIOT H. RODNICK
University of California, Los Angeles

NEVITT SANFORD
University of California, Berkeley

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
University of Pennsylvania

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The subscription price per volume is $10.00, or $20.00 annually (foreign, $20.50). Single issues are $4.00.

Original manuscripts should be sent to the Editor-elect:

DANIEL KATZ
Department of Psychology
University of Michigan
Ann Arbor, Michigan

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1961 by the American Psychological Association, Inc.

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 189-200

"CREDULOUS" AND "SKEPTICAL" VIEWS OF HYPNOTIC PHENOMENA:

EXPERIMENTS ON ESTHESIA, HALLUCINATION, AND DELUSION[1]

J. P. SUTCLIFFE[2]
University of Sydney

THERE is agreement about the phenomena of hypnosis so long as terms like hallucination, delusion, and amnesia are used in a purely descriptive sense. Concerning the interpretation to be put upon the hypnotic subject's behavior and testimony about his subjective experiences—that is, concerning the nature of hypnotic phenomena—there is traditional disagreement.

Some assume a priori a distinction between trance behavior and simulation, and accept in good faith the subject's testimony concerning the occurrence of "genuine" hypnotic phenomena. They believe that hypnosis can produce aberrations in any given mental or behavioral function and transcendence of normal capacity. They assume that hypnotic suggestions are substitutable for real stimuli as conditions of perception, and consequently that hypnotic fantasy is akin in its sensory content to perception produced by parallel real stimuli. Erickson (1952) is representative of this "credulous" view.

The "skeptics" do not accept hypnotic behavior and testimony at face value, as no unequivocal criterion for distinguishing trance behavior and simulation has yet been found. Furthermore they note that at least some hypnotic subjects adopt devious and misleading means of achieving suggested effects and misreport their actual experience (Pattie, 1933). They leave it an open question whether any given trance performance may not be matched in a nontrance state. They accede that an hypnotic subject may act as if the world were as suggested, taking his cues for a successful performance from the hypnotist's explicit and implicit directions (Sarbin, 1950; White, 1941) and from the "demand characteristics" of the situation (Orne, 1959); and they allow that there may be differences of subjective experience in trance and nontrance states; but they doubt that hypnotic fantasies have the same sensory content as parallel perceptions of real stimuli.

The main points at issue may be summarized in two questions of fact: Is hypnotic behavior like voluntary behavior? Is it akin to an actor's performance of a role? or Is hypnotic behavior relative to a "subjective reality" the sensory content of which is identical in form with that produced by the stimuli of a parallel objective reality? These questions may be answered by comparing hypnotic behavior with normal waking behavior and with waking acting or simulation.

For experimental study the two viewpoints need to be formulated in relation to specific hypnotic phenomena. A major difference between them concerns the subject's perceptions, so attention can reasonably be restricted to the cognitive phenomena of hypnosis. These may be classified in pairs such as paresthesia and anesthesia, positive and negative hallucination, and the like, according to whether the subject is required to produce or inhibit an experience. It then remains to select given modalities of sensory experience. For illustration, the two views are stated in relation to hypnotic anesthesia and paresthesia for painful stimulation.

The credulous view is that with hypnotic suggestion of anesthesia the subject has no more sensory experience in the presence of painful stimulation than he would have in the absence of such stimulation. Similarly, with hypnotic suggestion of paresthesia the subject has sensory experience in the absence of painful stimulation exactly as in the case where such stimulation is present. On this view, hypnotic suggestions are substitutable for real stimuli as conditions of sensory experience.

The skeptic doubts that painful sensory content is eliminated by suggestions of hyp-

[1] Arrangements for preparation and publication of this material were made during the author's tenure of a Rockfeller Foundation Fellowship in Social Science.

[2] For their help in various ways, I wish to express my sincere appreciation to A. G. Hammer, R. W. White, Martin T. Orne, and M. Brewster Smith.

notic anesthesia and that it is produced in paresthesia. He holds that the subject may agree with the hypnotist's suggestions in acting as if the suggested state of affairs were so. There are two possibilities, however, concerning the subject's perceptions:

1. *Simulation.* Although the subject complies with the suggestions he knows to the contrary that he is or is not being subjected to painful stimulation. Typically he misreports his true experience. If this be so, the central problem of hypnosis concerns the subject's motivation for misreport as a function of the hypnotic relationship and its procedures. In other respects no distinctions would be drawn between hypnotic behavior, voluntary behavior, and role playing.

2. *Delusion.* In complying with the hypnotist's suggestions the subject in some sense does not know that the real state of affairs is otherwise than suggested. If the subject is receiving painful stimulation there can be no passive ignorance, and "not knowing" must mean an active denial of reality. This version of the skeptical view which allows that the hypnotized subject may be deluded about some features of the trance situation has affinities with the credulous view. It may, however, be held independently of those aspects of the credulous view which concern sensory content.

Methodological weaknesses render equivocal much of the available, clinical and experimental evidence on these issues (see Sutcliffe, 1960).[3] Evidence alleged to support the credulous view can be shown to be consistent with the skeptical view, and most points at issue remain unresolved. A necessary albeit insufficient condition for resolution is improvement in the design of experiments on hypnosis. This paper reports a series of experiments conducted in awareness of these difficulties: a discussion of the experimental design is followed by descriptions of experiments on esthesia, hallucination, and delusion that bear on the credulous and skeptical views.

Design of Experiments

General requirements for the design of experiments on hypnosis are discussed elsewhere (Sutcliffe, 1958, 1960). The basic design is set out in Table 1. The major comparisons bear upon the questions of fact. The control conditions constitute the "parallel objective reality" of the second question; while suggestions given in hypnotic trance, on the credulous view, set up a "subjective reality." These are referred to, respectively, as the C and the HIS conditions. Comparison of behaviors under these two conditions bears upon the second question of fact and thus enables a test of the credulous view. The two hypnotic conditions may be compared to distinguish the effects of hypnotic induction, HI, per se from the effects of experimental suggestions given to the subject following upon the trance induction procedures, HIS. In the WA conditions the subjects are required to act in the waking state as if conditions were as suggested by the experimenter; and comparison of this behavior with that found in the other conditions enables a test of the skeptical view.

The presence and absence of stimuli refer to the nature of the real situation. Fantasy adds something that does not exist or affirms something false, or it takes away something that does exist or denies something true. One may thus denote absence in a real situation as the positive case, and presence in a real situation as the negative case. For example, C+ refers to the control condition with the experimental stimuli absent. To test the credulous view, one compares behavior under positive hypnotic fantasy, HIS+, with behavior under negative control conditions, C−: in the latter the stimulus is present, in the former the stimulus is actually absent but the hypnotized

TABLE 1
BASIC DESIGN FOR EXPERIMENTS ON HYPNOSIS

Major Comparisons	Stimuli	Hypnotizability	
		Somnambule	Nonsomnambule
Control (C)	Present	C−	C−
	Absent	C+	C+
Hypnotic induction (HI)	Present	HI−	HI−
	Absent	HI+	HI+
Hypnotic induction and suggestion (HIS)	Present	HIS−	HIS−
	Absent	HIS+	HIS+
Acting when awake (WA)	Present	WA−	WA−
	Absent	WA+	WA+

[3] Orne's (1959) work is a notable exception. It was not published, however, at the time the studies reported in this paper were conducted.

subject has to fantasy its presence. Conversely, one compares negative hypnotic fantasy with positive control behavior. If behaviors are alike in these comparisons the credulous view is confirmed; but if they are unlike, and especially if HIS+ is like C+, and HIS− is like C−, real stimuli are dominant and the skeptical view is confirmed. For these various comparisons to have any force, it is of course necessary that behaviors differ in the presence and absence of the control stimuli.

The experimental comparisons are made for both somnambulistic and nonsomnambulistic subjects. This provision rules out the possibility that failure to confirm the credulous view in certain cases is due to the use of insusceptible subjects. The complete design thus entails eight distinct experimental conditions replicated for two types of subjects. No one subject performs under more than one condition in a given experiment, that is, an experiment entails 16 independent groups of subjects.

Experiments on esthesia, hallucination, and delusion were conducted in conformity with this basic design. Each experiment entailed three main observational stages: pretesting for all three spanned 2 months and preceded major experimentation for all three which spanned 3 months before final subsidiary experimentation. The three experiments were thus conducted at the same time. They drew upon a common pool of subjects, but these were allocated among the experimental groups in different ways from experiment to experiment. A typical subject, for instance, could be a control in one case, a hypnotic subject in another, and an actor in the third.

Sampling and Allocation of Subjects

The experimental subjects were drawn from undergraduate psychology classes at the University of Sydney. Such students are highly selected from the general population in age, ability, and education. For this group women tend to outnumber men, and interests tend more to the humanistic than to the natural scientific.

For the experiments, two pools of subjects were to be obtained from this population: somnambules and nonsomnambules. These were obtained by selection from volunteers. The major criterion for selection was reaction to hypnosis. Subjects were handled individually, the initial session spanning 1.5 to 2 hours. Adapted from Bernheim (1890), Davis and Husband (1931), Friedlander and Sarbin (1938), White (1941), and Erickson (1952), the particular criteria were as follows: for *somnambules* manifestations on the first occurrence of hypnotic trance and at subsequent sessions of complete spontaneous posthypnotic amnesia; manifestations on the first or at subsequent sessions of "convincing" esthesias, hallucinations, and delusions, and of the ability to open the eyes and converse without coming out of trance; positive motivation; and for *nonsomnambules* reality contact and self-awareness throughout all hypnotic procedures with complete absence of spontaneous and suggested amnesia; compliant or willed rather than compulsive or involuntary reactions to suggestion; positive motivation.

Erickson's (1952) claim that success in hypnotic experimentation cannot be expected unless subjects are "trained" for deep trance is contentious; but to ensure that conditions were not unfavorable for a test of the credulous view, all somnambules were trained by the methods he describes to ensure clear-cut clinical manifestations of hypnotic esthesia, hallucination, and delusion prior to the beginning of experimentation. There was a minimum interval of 2 months between selection and training on the one hand and participation in an experiment on the other. During the experiments proper which involved hypnosis, a subject was run through a trance induction procedure that took 30 minutes. For a subject chosen for the HIS conditions, this was followed by an additional 15 minutes of hypnotic procedure for the production of esthesia, hallucination, or delusion as required (see Sutcliffe, 1958, pp. 105–108, 219–249, for details).

Of 487 students approached, hypnotic procedures were applied to 156 volunteers; and of these, 108 were rejected as recalcitrant or as intermediate or uncertain in their degree of hypnotizability. Selection continued as subjects came to hand until 24 somnambules and 24 nonsomnambules had been found, each pool comprising 12 men and 12 women. There were no significant differences in age for these four subgroups, and the overall mean was 22.9 years with standard deviation 3.2. For a given

type the incidence out of 156 is approximately 16%, which agrees with the statistics compiled by Bramwell (1906). Thus although the specific aim in selection was to obtain given numbers of subjects, it happened that the two extreme types of subjects occurred at about the same rate as they do for the general population insofar as that can be determined from inadequate statistics; and so the experimental universe is not atypical. Experimental results obtained from the samples have generality in that the latter are at least representative of the extremes of hypnotizability for young adult men and women at the university.

The primary concern in testing the credulous and skeptical views is to compare the control, hypnotic, and acting conditions; studies of the effects of trance per se, of individual differences and reliability are subsidiary. Attention might thus be restricted to methods of allocating subjects to the six experimental conditions: C+, C−, HIS+, HIS−, WA+, and WA−. For the experiment on esthesia, six independent random samples of four subjects were drawn separately from each of the somnambule and nonsomnambule pools, and one of the six experimental conditions was assigned at random to each group. With the other two experiments a matching procedure was used to minimize intersubject variation which, in samples as small as four, could blur group differences. To achieve this matching, subjects were listed in rank order on a relevant pretest, the parallel form of which was to be used as the dependent variable in an experiment. Twenty-four somnambules could thus be subdivided into six subgroups each containing four serially ordered subjects. The same procedure was applied to the nonsomnambules. Although the common pool of 48 subjects was used for all three experiments the experimental allocations for a given subject varied from experiment to experiment. No subject was aware of his group allocation for any given experiment, and his participation could give him knowledge of only 1 of 12 conditions in a given experiment. The experimental testing was embedded in a context of other testing procedures; and in these various ways all subjects were kept ignorant of the overall design of the experiment and of the experimenter's aims.

EXPERIMENT I: ESTHESIA

Esthesia was chosen for investigation as it is regularly cited in support of the credulous view. Most commonly, hypnotic anesthesia has been studied; here paresthesia is also considered. Most studies have used some sort of a probe as a pain stimulus; here electric shock is used. The subject's testimony, gross motor movements, and GSR to the shock stimuli are observed.

Method

Apparatus. The subject was connected to a GSR circuit (Woodworth, 1938, p. 278) with electrodes attached to fingers of the right hand, and to a dc shock circuit via the left hand. A switch in the latter circuit enabled the experimenter to deliver either a buzz without shock, or buzz and shock simultaneously.

Pre-experimental testing. The subjects were given a series of shocks with the apparatus described. This provided them with experience necessary for subsequent paresthesia if it were to occur. Also it enabled the experimenter to determine a fixed level of electric shock that produced pain in all subjects.

Experimental procedures. The experiment was conducted in a sound-dampened room. The subject sat at a small table placed in a corner so that ahead and to the right he faced bare walls, while to the left and behind he was bounded by a wooden screen. The experimenter sat behind the subject, or moved out when necessary to control the apparatus concealed by the screen. The table was bare and only the electrodes were visible. A fluorescent source provided constant illumination. The only stimuli of a variable character were the experimenter's instructions, and the buzz or buzz-shock of the apparatus as delivered by the experimenter.

The particular experimental conditions were as follows:

- C^+ The subject was linked to both the shock and GSR apparatus but received no shocks. When the stimulus key was depressed by the experimenter, buzz alone occurred. The only instructions given to the subject were that he should remain passive and refrain from gross motor movements.
- C^- This condition was identical with C^+ except that when the experimenter depressed the stimulus key, the subject received an electric shock simultaneously with hearing the buzz.
- HIS^+ The objective conditions were those of C^+. In addition, the subject was induced into deep trance and paresthesia for shock suggested, the "shock" to be felt simultaneously with hearing the buzz.
- HIS^- The objective conditions were of C^-. In addition, the subject was induced into deep trance and anesthesia was suggested for the left hand.
- WA^+ With the objective conditions of C^+ the subject was asked to recall a circumstance in

which, though not in pain, he had behaved as if he were, specifically, touching an electric iron to see whether it was hot; and the subject was then asked to respond at the buzz as if he had received a painful electric shock of the magnitude previously experienced.

WA⁻ With the objective conditions of C⁻ the subject was asked to recall a circumstance in which, though in pain, he had behaved as if he were not; specifically, having fallen over as a child when playing with other children; and the subject was then asked to respond at the buzz as if there were no shock.

In each case, the subject received four stimuli at intervals of approximately 1.5 minutes. In the intervals the experimenter recorded the subject's GSR and the base level of skin resistance prior to giving the next stimulus, and he repeated crucial instructions pertaining to the experimental condition. Hypnotized subjects were specifically questioned about their experiences. Movements and comments from subjects provided qualitative data to supplement the quantitative GSR records. The room was sound dampened, the experimenter worked quietly, the subject remained passive, thus irrelevant stimuli to the GSR were absent, and the GSR can be taken as indicative of the subject's reaction to the experimental condition.

Predictions

On the credulous view, one would expect the GSR for hypnotic paresthesia to be the same as the GSR for the C⁻ condition: that is, fantasied shock should be equivalent to experience of real shock. Also, GSR for hypnotic anesthesia should not differ from GSR in the C⁺ condition if fantasied absence of shock is equivalent to actual absence. At the same time it would be expected that waking subjects cannot imitate the hypnotic performance. Furthermore, while nonsomnambules may be like somnambules in the control and acting conditions they should differ from them in the hypnotic conditions. Complete failure of these outcomes would confirm the skeptical view.

Results

Qualitative data are reported in Tables 2 and 3 for the eight somnambules who were tested under the HIS+ and HIS− conditions. These describe movements and comments made by the subject as trance reactions and in a comparable waking condition. These data show that as far as overt signs and the subject's testimony go, hypnotic paresthesia and anesthesia were invariably experienced.

GSR data are presented as log conductance change (after Haggard, 1949). If R is a base level of skin resistance immediately prior to stimulation, and R′ is level of resistance at the peak of reaction to the stimulus, the measure of reaction used is $(40 + 10 \log - 1/R' - 1/R)$. This scaling with log zero denoted bar four gives a range of values from 0 (no reaction) to 30, the largest GSR for these data. Mean values for the various conditions are set out in Table 4.

The data were tested for significance by F and t tests via the analysis of variance, and the main findings were: Measurement of the GSR was reliable, intersubject variation being significantly ($p < .001$) greater than intrasubject variation. There was a significant

TABLE 2

QUALITATIVE OBSERVATIONS ON REACTIONS TO REAL SHOCK (Pre-experimental) AND TO HYPNOTIC PARESTHESIA OF SHOCK

Subjects	Subject's Reactions with Paresthesia	Subject's Waking Reactions to Real Shock
1	Flinches, grimaces, moves hand. In reply to questions about what he felt, the subject says dislikes shocks and would prefer that no more be given. Needs reassurance to continue.	Most apprehensive of apparatus; needs reassurance of low dc amps before proceeding; shakes electrodes off fingers at each shock. Has to be encouraged to replace electrodes and to continue.
2	No gross movements. When questioned, the subject reports that he feels strong shocks.	No gross movements apart from finger reaction but says with vehemence: "Wow! that gave me a belt."
3	Minor hand movement, facial grimace. Says first shock was only mild. Says that subsequent shocks were getting stronger and not to increase the strength further.	Reluctant to participate, says, "I know it's silly but I hate shocks." Needs much reassurance before proceeding. Exclaims loudly and shakes hand at each shock. Strong anticipatory avoidance both in hand movement and in vocal "oo."
4	Movement of hand; otherwise no signs. When questioned says she feels shocks but can take them if she has to.	Stoic, calm, and, apart from finger movement, no overt sign of pain. When questioned the subject reports that shocks were strong, can take them, but would not like them any stronger.

TABLE 3
QUALITATIVE OBSERVATIONS ON REACTIONS TO ACTUAL ABSENCE OF SHOCK (Post-experimental) AND TO HYPNOTIC ANESTHESIA TO SHOCK

Subjects	Subject's Reactions with Anesthesia	Subject's Waking Reactions Without Shock
1	Slight hand movement but reports feeling nothing. Visible increase in respiration subsequent to shock.	Keyed up expecting shock; exhibits startle at "buzz" and relaxes when he finds he is not shocked; thereafter passive.
2	Stuporous passivity; no movement in reaction to shock. When questioned, has to be urged to reply and mumbles that he feels nothing: "Hand is stone."	Apprehensive of shock; but reassured when only "buzz" occurs. Startled at first stimulus, but thereafter passive.
3	Stuporous passivity; no movement apart from hand jerk to shock. When questioned, denies feeling anything.	Says, "I hope there won't be any shocks." Some startle in response to "buzz," says, "That didn't hurt" and laughs. Subsequently passive.
4	Whole body jumped. When questioned the subject said, "I don't feel anything, but *she* seems uncomfortable." Subsequent movement restricted to fingers, but denies any feeling.	Matter of fact approach to the situation; no outward reactions. After second stimulus asks whether she should be receiving shocks. Accepts avoidance of reply with equanimity.

TABLE 4
MEAN CHANGES IN GSR EXPRESSED AS LOG CONDUCTANCE CHANGE

	Somnambule		Nonsomnambule	
	Nonshock	Shock	Nonshock	Shock
Control	3.3	18.1	10.8	19.7
Hypnosis	8.7	19.4	9.1	18.9
Acting	16.6	18.1	13.3	18.5

($p < .0001$) overall mean difference of 8 reaction units between the shocked and the nonshocked subjects, with all shock groups averaging about 18 to 19 while the nonshock groups differed widely. In the nonshock conditions, hypnotic subjects gave reactions consistent with the absence of shock, whereas the actors gave greater reactions approaching in magnitude those given by shocked subjects. That is, neither hypnotic anesthesia nor waking acting was able to inhibit shock reaction and in this they were alike; while acting could produce a shock-like reaction and hypnotic paresthesia did not and in this they were not alike. In reaction through GSR the somnambules and nonsomnambules were similar. In the comparisons that bear most directly upon the credulous and skeptical views, HIS+ was much more like C+ than like C−, and HIS− was much more like C− than like C+, and in this the data do not support the credulous view.

To check on the effects of trance induction per se, the eight somnambules of the C+ and C− groups were subsequently hypnotized, and with this difference the previous control procedure was repeated. The two groups, which averaged 3.3 and 18.1 in the waking state, average 3.5 and 17.4, respectively, in the trance state, so that the sheer fact of being in trance did not affect the GSR.

In case intersubject variability contributed to the nonsignificance of certain differences, given conditions were compared using each subject as his own control. The 16 somnambules comprising the HIS+, HIS−, WA+, and WA− groups were retested under the C+ and C− conditions; and thus a given subject's experimental performance could be compared with his normal waking reactions in the presence and absence of shock. These results fully confirmed those obtained with the independent group comparisons.

Conclusions

The credulous view that hypnotic suggestion is substitutable for real states of affairs as conditions of sensory content is clearly refuted by the quantitative data. The real state of affairs, shock or its absence, is the determinant of reaction even in the face of contradictory fantasy. It is quite clear that hypnotic suggestion does not produce "anesthesia" if by this is meant "absence of GSR to electric shock"; nor does it produce "paresthesia" if by this is meant "occurrence of shock-appropriate GSR without electric shock." On the other hand, problems remain when these

findings are considered in relation to the subject's testimony. Hypnotically anesthetic subjects report that they feel no pain when given shocks known otherwise to be painful and when their bodies are definitely reacting to the stimuli. Hypnotically paresthetic subjects report that they feel pain despite the absence of shock and bodily reaction. As reported, subjective experience is at variance with bodily reaction.

EXPERIMENT II: HALLUCINATION

To investigate hypnotic hallucination—central to the credulous view—the auditory modality was chosen. Subjects were required to hallucinate, positively or negatively, sounds known to have a measurably distracting effect upon performance. This would provide an objective check upon the subject's testimony about hallucination.

Method

Apparatus. It is known that delayed feedback of speech can disrupt and slow down speech (Spilka, 1954; Tiffany & Harley, 1952) and this fact was utilized in the experiment. The method used to produce the effect was to have the subject speak into a microphone attached to a multihead magnetic tape recorder arranged so that the input was fed into the subject's earphones after a delay of .27 seconds. (The reported optimal delay of .20 seconds could not be set up as the machine available provided only for delays of .07, .13, and .27 seconds.) The subject's vocalizing was an integral part of his test performance which then became subject to distraction effects.

The test performance required the subject to say aloud each step and the answer to simple addition problems. An item comprised the addition of three two-place numbers; there were 20 items to a trial, and five trials to a form. The form was prepared as a printed sheet of five rows of addition items to be presented to the subject for working and recording his answers. A 50-second time limit was set for each trial, and within this time the average subject could complete 10 items correctly. Parallel Forms A and B were prepared with constant mean and variance and estimated reliability greater than .8.

In preliminary standardization of the delay technique, there were observed progressive adaptation effects with practice and individual differences in magnitude of distraction effect positively correlated with speech rate. By comparison with normal average performance on the addition test, however, subjects performed significantly less well with delayed speech feedback.

Preliminary testing. Subjects were tested on Form B of the addition test under normal conditions. This provided a criterion for matching subjects for the experiment to follow; it familiarized the subjects with the test procedure so that detailed explanations could subsequently be avoided, especially in the experimental hypnotic sessions; it provided baseline data for assessing the effects of trance per se on test performance; and it provided sensory experience for subsequent hallucination should it occur.

Experimental procedures. The experimental setting was similar to that described for the experiment on esthesia. The table was bare except for the test form and pencil which the experimenter provided when the test was to begin. Whether subjected to delayed feedback or not, every subject wore earphones fitted tightly to minimize normal air conduction of sound. The experimenter's instructions and the subject's vocalizings were routed to the earphones through a microphone located to the left front of the subject at a maximum distance of 18 inches from his mouth.

The particular experimental conditions were as follows:

- C^+ In the waking state the subject performed on Form A of the addition test. He wore the earphones and the recorder was set to produce zero delay. The subject did not speak aloud but mouthed words exactly as when speaking except for the absence of sound.
- C^- In the waking state the subject performed on Form A. He wore the earphones and the recorder was set to provide .27 seconds feedback delay. The subject spoke aloud as he worked at the test and suffered distraction.
- HIS^+ The objective conditions were those for C^+. In addition, the subject was induced into deep trance and positive hallucination for his own voice was suggested. The experimenter illustrated the interval of delay at which the subject was to hear his voice whenever he mouthed words, and the procedure was repeated until correct.
- HIS^- The objective conditions were those for C^-. In addition, the subject was induced into deep trance and negative hallucination for his own voice was suggested.
- WA^+ The objective conditions were those for C^+. In addition, the subject was asked to recall a circumstance where he behaved as if he had heard a noise when there was in fact none, specifically, querying a person who appeared to have spoken; the experimenter illustrated the interval of delay at which the subject was to imagine hearing his voice whenever he mouthed words, repeating the procedure until correct; and the subject was then asked to perform the task as if, while mouthing words, he were actually hearing his own voice slightly delayed.
- WA^- The objective conditions were those for C^-. In addition, the subject was asked to recall a circumstance where he behaved as if he had not heard some sound, specifically, by ignoring a call when engrossed in an activity; and the subject was then asked to perform the task as if he could not hear his own voice.

In each case, the subject performed five trials at intervals of approximately 30 seconds. Crucial instruc-

tions were repeated between trials. Qualitative observations were restricted to the subject's testimony concerning the presence of hallucination. Quantitative data comprise the five trial scores for each subject.

Predictions

On the credulous view one would expect the performance of subjects with positive auditory hallucination to be like that for the negative control condition; that is, fantasied distraction is supposedly like real distraction. Also, performance under HIS− conditions should be like that for the positive controls if negative auditory hallucination does in fact offset actual delayed feedback distraction. At the same time it would be expected that waking actors cannot imitate these performances. Somnambules would be expected to differ from nonsomnambules under HIS conditions. Complete failure of these outcomes would confirm the skeptical view.

Results

The qualitative data for the hypnotic conditions were as follows. With the somnambules, testing was not begun until appropriate hallucination was reported; affirmative testimony was given in all cases. With the nonsomnambules, in no case did a subject accept the suggestions.

The addition test data are presented as number of correct items per trial with a possible range from 0 to 20. Mean values for the various groups are set out in Table 5. The results for the 12 groups on the Form B pretest under normal conditions are set out on the left-hand side; and their results on Form A under experimental conditions are set out on the right. These data were tested for significance via analysis of variance. There is a small but significant pretest difference in performance between the somnambules and nonsomnambules which does not affect the main outcomes of the experiment proper. Test performance is reliable, intersubject variation being significantly ($p < .001$) greater than intrasubject variation. The only significant effect for the experimental data ($p < .05$) is the overall mean difference of approximately two items per trial between presence and absence of delayed speech feedback distraction. The data for the nonsomnambules are accounted for in this way: in the absence of distraction the pretest and experimental performances are the same; in the presence of distraction performance drops from the pretest level by an average of approximately two items per trial. The data for the somnambules in the HIS conditions cannot be explained simply in terms of distraction, and account must be taken of the effects of hypnosis per se.

TABLE 5
MEAN PERFORMANCE ON ADDITION TEST FORMS A AND B

	Form B, Pretest				Form A, Experimental			
	Somnambule		Nonsomnambule		Somnambule Distraction		Nonsomnambule Distraction	
	Normal Testing Conditions for All Groups				Absent	Present	Absent	Present
Control	10.8	11.0	9.2	9.0	10.8	8.2	9.3	7.3
Hypnosis	11.0	10.5	9.2	9.5	8.5	7.0	9.6	7.4
Acting	10.2	11.5	9.2	9.2	10.0	9.0	9.1	7.0

The 16 somnambules of the C+, C−, WA+, and WA− groups were retested with Form B when later induced into trance. Comparison of individual results for the two testings revealed uniformly poorer performance in trance, the average drop exceeding two items per trial. Whether lethargic performance is an intrinsic feature of deep trance or a specific effect of the sleep suggestion procedure for inducing trance in these experiments is not established. The drop due to trance lethargy must be allowed for in this instance; together with the effects of delayed feedback distraction it accounts for the results of the somnambules in the HIS groups. The drop from 11.0 to 8.5 for the HIS+ group is explicable in terms of trance lethargy rather than positive hallucination of distraction; while a combination of trance lethargy and actual distraction can account for the drop from 10.5 to 7.0 for the HIS− group.

The eight somnambules of the HIS+ and HIS− groups were retested with Form A when in trance but without suggestions of hallucination. This procedure enabled comparison for a given subject of performance of the addition test under the three conditions: waking, trance, and trance plus experimental suggestions. These data, with intersubject variability controlled, corroborated the explanation of the results for the independent groups in terms of trance lethargy and speech feedback distraction.

Conclusions

Waking acting and hypnotic "hallucination" are alike in this case in that they are equally unsuccessful in opposing the effects of real stimuli. As with the experiment on esthesia, it is the real state of affairs rather than the subjective experience that determines the subject's objective behavior. Hypnotic suggestion does not produce positive hallucination if by this is meant sensory content comparable to that produced by actual speech feedback delay; neither does it produce negative hallucination if by this is meant elimination of distraction effects. The credulous view is clearly refuted by these data. There remains, however, the puzzling discrepancy between the subject's reports of hallucination and the evidence of their objective performance.

EXPERIMENT III: DELUSION

As a major category of fantasy, delusion recommends itself for study; but there is a dearth of experimental evidence. Delusion is typically defined as a false belief resistant to modification by contrary evidence. Thus a prerequisite for study is the choice of a proposition for belief. Of the infinite number available it is convenient to choose one the subject of which is the person or his ego: then the delusion has immediate and wide implications for the subject's behavior. Thus something is to be falsely predicated of the subject's ego. In terms of the design, the proposition must take both a positive and a negative form. The predicate must thus be such that it can be truthfully affirmed of one group and falsely affirmed of another, and at the same time, it can be truthfully denied of one group and falsely denied of another. As the experimental groups to be used contained equal numbers of men and women, "sex" was a predicate that fulfilled all of these conditions. The logical form of the positive and negative propositions for men and women is given as:

		MAN	WOMAN
TRUE	+	You are a man	You are a woman
	−	You are not a woman	You are not a man
FALSE	+	You are a woman	You are a man
	−	You are not a man	You are not a woman

The positive and negative conditions are different in that the statements that form a positive-negative pair are not equivalent. The positive statement is unambiguous, but the negative allows several interpretations: for example, "You are not a man" need not imply "You are a woman"; the subject's interpretation might be "I am a child." These direct statements were not used; the intention of the logical form was achieved by more elaborate suggestions. The details are given in Sutcliffe (1958).

The experimental test of the credulous view of delusion involves the subject in behavior relative to a belief about his sex. Of the many ways of behaving characteristic of the sexes, moral and ethical considerations ruled out direct observation of the more obvious genital behaviors. Interrogation relative to "hypothetical situations" might have yielded information, but this would have been limited to the subject's testimony. Accordingly attitudes and interests characteristic of the sexes were investigated.

Method

Materials. Two parallel Forms A and B, each of 92 items, were prepared from the Emotions and Interests sections of the Terman and Miles (1936) Attitude-Interest Inventory. An item is scored +1 for a masculine response, −1 for a feminine response, and 0 for a neutral response by reference to the Terman-Miles scoring key. The algebraic sum of the weights gives a subject's score for the form. The correlation between the forms is .81 under normal conditions. The average score for women on the forms is just less than 0, $SD = 13$. The average for men is about 18 points higher, $SD = 10$. There is little overlap of the distributions for men and women; but it is known from the Terman-Miles study that a subject can, upon request, make answers conform or overconform to those of the opposite sex.

Preliminary testing. All subjects were pretested on Forms A and B of the Attitude-Interest Inventory. This provided a criterion, Form B, for matching subjects in the experiment to follow, and a baseline, Form A, from which change in the experiment could be measured. While subjects would be most familiar with their own sex roles, they would also have knowledge of opposite sex roles, and accordingly no special experience had to be provided to enable subsequent delusion were it to occur.

Experimental procedures. The setting was that described for the experiments on esthesia and hallucination. The table was bare except for the test Form A and pencil which the experimenter provided at the appropriate time. When the subject had completed the inventory, the experimenter questioned the subject about his responses and sexual identity, and these data were recorded on tape.

The particular experimental conditions were as follows:

C^+ Half the subjects were men and for them it was

simply true that they would fill out the inventory as men. The other half were women and would fill out the inventory as such. Each subject was required to enter name, age, and sex at the beginning of the test form before proceeding.

C⁻ Half the subjects were women and hence not men; while the other half were men and hence not women. The manner of filling out the inventory was as in the C⁺ condition.

HIS⁺ The objective conditions were those of C⁺. In addition, the subject was induced into deep trance when a delusion about sex was suggested in line with the logical form: "You are a woman" for a man; and "You are a man" for a woman.

HIS⁻ The objective conditions were those of C⁻. In addition, the subject was induced into deep trance when a delusion about sex was suggested in line with the logical form: "You are not a man" for a man; and "You are not a woman" for a woman.

WA⁺ The objective conditions were those of C⁺. In addition, the subject was asked to recall a circumstance where he (she) played the part of the opposite sex; and the subject was then asked to fill out the inventory as if he were a woman (as if she were a man).

WA⁻ The objective conditions were those of C⁻. In addition, the subject was asked to recall a circumstance in which he (she) had denied his (her) sex; and the subject was then asked to fill out the inventory as if he were not a man (as if she were not a woman).

Predictions

Whereas with the experiments on esthesia and hallucination it was not known in advance whether the subjects could perform the tasks set, in the experiment on delusion it was known that a waking subject can to some extent simulate opposite sex responses on the Attitude-Interest Inventory. Thus instead of asking whether an actor may do as well as a hypnotized subject, one must ask whether they differ in any way. On the credulous view, a deluded subject should behave just as a subject who holds the same belief in truth. Thus not only would one expect a change in performance with a suggested change in sex, but also it should be appropriate in degree. While alike in the content of their beliefs, however, the two groups may well differ in their reactions to interrogation and contrary evidence. Somnambules and nonsomnambules would be expected to differ primarily in the hypnotic conditions.

Results

The interrogation sought answers to two questions: what does the subject believe his sex to be? and how would he prove this belief? The details of the questioning and the answers given varied from case to case, but the results may be summarized as follows:

All control subjects correctly recorded their sex, and when questioned about it they invariably cited biological sexual characteristics as proof. Less frequently cited in addition were the judgment of others, birth certificate, clothing and hair styles, and interests. Subjects showed that they regarded their sex as self-evident, and they were puzzled by the experimenter's questioning of the obvious.

All "acting" subjects correctly recorded their sex, and they cited supporting evidence of the same kind as that given by the controls. For all but one subject, a nonsomnambule, the act terminated with the completion of the inventory. This subject carried the act into the interrogation, but when pressed for evidence gave up the act and changed to an admission of his correct sex. In no case was the act taken as far as entering the incorrect sex at the head of the inventory. When asked about this, subjects gave the stock reply that this was identifying information which stood apart from the experimental context. Two nonsomnambules and four somnambules reported difficulty with acting, and one of the latter said that she tried to give answers typical of neither men nor women.

All nonsomnambules in the hypnotic conditions correctly recorded their sex and gave appropriate evidence in support. All reported difficulty in imagining the state of affairs suggested by the hypnotist, and two of the men took objection to the suggestion that they had become women and actively opposed it.

The somnambules in the hypnotic conditions differ from the other subjects in several ways: While all 40 subjects so far considered correctly recorded their sex at the head of the form, only 2 of these remaining 8 subjects did so, one woman entered "male," and 5 made no entry. The 4 subjects in the HIS+ group, though in deep trance, resisted the suggestion that they were of the opposite sex and passively but continuously reiterated their true sex. This behavior was unlike that of the controls in its repetition and failure to mention any evidence. While the 4 somnambule and 2 nonsomnambule subjects of the HIS+ condition objected to the suggestion that they were of the opposite sex, the 4 somnambules of the HIS− group more readily accepted the denial of their sex. They made statements such as "I feel it quite strongly, I'm sure of it" and "I feel it's just obvious"; and in this their

conviction about the assumed sexual identity was like the controls' conviction about real sexual identity. They did not however cite any evidence.

Quantitative data were analyzed for change as follows. The difference in score between pretest and experiment on Form A was found for each subject. The average change for a control group was subtracted from each subject's change score. This correction was carried out separately for the somnambule and nonsomnambule groups as the two sets of control groups differed in retest variation. The difference in direction of change due to sex was corrected by denoting change scores positive if the change was in the suggested direction, and negative if the change was in the contrasuggested direction. It will be seen that this procedure sets the change scores for the control groups to zero. The mean values of change scores for the eight experimental groups are set out in Table 6. Differences between negative and zero, and between zero and positive change are significant.

Under normal conditions there is a mean difference of 18.6 between men and women on Form A. Thus for a given sex to successfully simulate the performance of the opposite sex group, it must change from its normal mean by 18.6. This provides a criterion for evaluating the data of Table 6. The nonsomnambules in the two hypnotic conditions did not change at all, consistent with their insusceptibility to hypnosis. On the other hand the nonsomnambules, when asked to simulate in the waking state, were able to change their performance and if anything they overacted the part. The somnambules in the HIS+ condition changed somewhat but inappropriately; the overemphasis in test performance upon true sex was consistent with their subsequent reaction to interrogation. The somnambules in the HIS− and WA− conditions changed by about the predicted amount, while those in the WA+ group overacted. Independent checks showed that trance per se does not affect performance on the inventory; and overall the results held for individual subjects as well as for the group averages.

TABLE 6
MEANS OF CHANGE SCORES ON THE ATTITUDE-INTEREST INVENTORY FORM A FOR THE HYPNOSIS AND ACTING GROUPS

Condition	Somnambule	Nonsomnambule
HIS+	−9.5	0.8
HIS−	18.8	0.3
WA+	26.8	27.0
WA−	15.8	25.2

Conclusions

Negative hypnotic delusion was successfully suggested to the somnambules as indicated by their testimony and by the appropriate mean change in performance. The WA− actors were as successful in changing their test performance, but their reactions to interrogation differed from those of the hypnotized subjects. In the HIS+ condition, both somnambules and nonsomnambules found the suggestion that they were of the opposite sex objectionable; the suggestion was resisted, the experimental conditions for the test of the credulous view were not fulfilled, and hence this part of the data must be set aside. It is of interest to note that deep trance does not guarantee acceptance of suggestion in all cases. Where the experimental conditions were met, the quantitative data do not clearly differentiate between the credulous and skeptical views; the qualitative data support the view that in a number of cases hypnotized subjects were deluded about their sex.

GENERAL SUMMARY AND CONCLUSIONS

Evidence from the three experiments shows that:

1. It was not possible by hypnosis to produce sensory conditions identical in form with those produced by parallel real stimuli. The GSR to real shock occurred despite hypnotic "anesthesia," and there was no GSR in the absence of shock despite hypnotic "paresthesia." Decrement in performance due to actual delay of speech feedback occurred despite hypnotic "negative hallucination," and no decrement occurred in the absence of delay despite hypnotic "positive hallucination."

2. There is equivocal support for the view that hypnotic behavior is like waking acting. The two were alike in that neither was successful in inhibiting the effects of shock or distraction, and a further similarity was found with the negative delusion about sex. The similarity in the first two cases should not be emphasized as it depends upon a failure of inhibition where real states of affairs are dominant. In

the positive condition where production rather than inhibition was called for, differences between hypnosis and acting did occur. In the absence of shock hypnotic subjects reported subjective experiences of pain but gave no GSR; while the actors with no subjective experience of pain overtly acted as if shocked and produced GSR.

3. With the exception of the HIS+ group in the third experiment, all hypnotized subjects were convinced of the actuality of suggested states of affairs. They asserted that they were shocked when in fact there was no shock, or they were not shocked when in fact they were; they could hear their own voices when they were not in fact speaking aloud, or they could not hear their own voices when in fact they were speaking aloud; they were not men (not women) when they were in fact men (women). This testimony was uniformly at variance with the evidence of the subjects' bodily states.

These findings justify the following conclusions:

1. The credulous view that hypnotic suggestions are substitutable for real states of affairs as conditions of sensory experience must be rejected.

2. The skeptical view that hypnotic behavior is like waking acting is true in some cases. The difference between hypnosis and acting might best be sought elsewhere than in the overt behavior, which is sometimes similar and sometimes dissimilar for the two depending upon the task and whether inhibition or production is called for. The qualitative evidence suggests that differences might more fruitfully be sought in terms of perception and belief.

3. Crucial evidence is lacking to decide between the two versions of the skeptical view concerning the subject's perceptions: that the hypnotic subject does not misperceive the real situation but misreports it; or that the hypnotic subject misperceives, that is to say is deluded about, the real situation. The latter view gets some support from the different reactions of hypnotic, control, and acting subjects to interrogation.

4. The distinguishing feature of hypnosis appears to be the subjective state; and the main feature of this state is the hypnotized subject's emotional conviction that the world is as suggested by the hypnotist, rather than a pseudoperception of the suggested world.

The role of the hypnotic relationship and its procedures in the communication of delusory conviction remains to be clarified. An approach which at the same time may throw some light on individual differences in hypnotizability would be to examine the authority relationship and the conditions of credence in their developmental aspects with emphasis upon the parent-child relation.

REFERENCES

BERNHEIM, H. *Suggestive therapeutics.* (2nd ed.) (Trans. by C. A. Herter) Edinburgh & London: Young J. Pentland, 1890.

BRAMWELL, J. M. *Hypnotism—Its history, theory and practice.* (2nd ed.) London: Alexander Moring, 1906.

DAVIS, L. W., & HUSBAND, R. W. A study of hypnotic suggestibility in relation to personality traits. *J. abnorm. soc. Psychol.*, 1931, **26**, 175–182.

ERICKSON, M. H. Deep hypnosis and its induction. In L. LeCron (Ed.), *Experimental hypnosis.* New York: Macmillan, 1952. Pp. 70–112.

FRIEDLANDER, J. W., & SARBIN, T. R. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 453–475.

HAGGARD, E. A. On the application of the analysis of variance to GSR data: I. The selection of an appropriate measure. *J. exp. Psychol.*, 1949, **39**, 378–392.

ORNE, M. T. The nature of hypnosis: Artifact and essence. *J. abnorm. soc. Psychol.*, 1959, **58**, 277–299.

PATTIE, F. A. A report on an attempt to produce unilocular blindness by hypnotic suggestion. *Brit. J. med. Psychol.*, 1933, **15**, 230–241.

SARBIN, T. R. Contributions to role-taking theory: I. Hypnotic behavior. *Psychol. Rev.*, 1950, **57**, 255–270.

SPILKA, B. Some vocal effects of different reading passages and time delays in speech feedback. *J. speech hear. Dis.*, 1954, **19**, 37–47.

SUTCLIFFE, J. P. *Hypnotic behaviour: Fantasy or simulation.* Unpublished PhD thesis, University of Sydney, 1958.

SUTCLIFFE, J. P. "Credulous" and "sceptical" views of hypnotic phenomena: A review of certain evidence and methodology. *Int. J. clin. exp. Hypnosis*, 1960, **8**, 73–101.

TERMAN, L. M., & MILES, C. C. *Sex and personality.* (1st ed.) New York: McGraw-Hill, 1936.

TIFFANY, W. B., & HARLEY, C. N. Delayed speech feedback as a test for auditory malingering. *Science*, 1952, **115**, 59–60.

WHITE, R. W. A preface to the theory of hypnotism. *J. abnorm. soc. Psychol.*, 1941, **36**, 477–505.

WOODWORTH, R. S. *Experimental psychology.* New York: Holt, 1938.

(Received February 25, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 201-209

PERFORMANCE ON COMPLEX REASONING TASKS AS A FUNCTION OF ANXIETY[1]

CHARLES VAN BUSKIRK[2]
University of Chicago

A good deal of recent research has been done concerning the effects of anxiety considered as an enduring personality trait upon performance in learning and problem solving. Considerable evidence has shown that high anxious (HA) groups are inferior to low anxious (LA) groups in their performance on complex tasks. The findings have suggested that a general relationship exists between the complexity of the task and the magnitude of the decrement produced by anxiety. Farber (1955) suggested the following proposition: "... the quality of performance in complex learning situations is inversely related to *Ss*' degree of anxiety, as measured by the Taylor scale, and furthermore, that the advantage of the nonanxious over the anxious *Ss* is positively related to the probable number and strength of the competing responses elicited" (p. 323). If a test, such as Taylor's (1953) Manifest Anxiety (*MA*) scale, is assumed to provide a measure of general drive level, the interaction between anxiety and certain task characteristics is predicted by Hull's hypothesis that response strength is a multiplicative function of habit strength and drive level. Poorer performance by the HA group is expected, for example, in either paired associate or serial learning tasks in which pre-existing associative relationships among the task elements lead to wrong responses, since the subjects with higher drive will take longer to extinguish their stronger erroneous response tendencies (Farber & Spence, 1953; Ramond, 1953; Spence, Farber, & McFann, 1956; Taylor & Spence, 1952).

It has been suggested that HA subjects make more irrelevant responses to a task—particularly a stressful or difficult task (Child, 1954). The drive theorist's position is that, if irrelevant responses are produced, they will compete with task appropriate responses and interfere with performance. Higher levels of drive will increase the number of irrelevant response tendencies above threshold at a given time and will decrease the rate at which any specific weak response can be selectively reinforced (Spence, 1958). A contrasting interpretation involves the direction of irrelevant responses. The essential point is made in Rosenzweig's (1944) distinction between "need persistent" and "ego defensive" reactions to frustration. In a need persistent reaction, behavior is oriented toward attaining an objective goal, that is, toward interaction with environmental conditions that are relevant to the frustration. In an ego defensive reaction, behavior involves a general avoidant or counteractive orientation which is irrelevant to the objective task. The latter characteristics have been attributed to anxious behavior by Hanfmann (1950) in her review of research on anxiety. She points out that anxiety is most commonly observed in an individual who is experiencing a threat that he does not clearly perceive or understand. Irrelevant ego defensive responses should be characteristic of HA subjects, since many of the items in the *MA* scale refer to a propensity to become anxious without regard for reality considerations. However, recent research has cast considerable doubt on the hypothesis that the HA subject produces more irrelevant responses than any other subject. On tasks which minimize associative interference, the performance of the high anxious group is frequently superior to the LA group (Spence et al., 1956; Taylor & Chapman, 1955). Even under some of the stress conditions which consistently produce performance decrements in subjects generally—very possibly by evoking irrelevant responses (cf. Lazarus, Deese, & Osler, 1952), HA groups have performed at a superior level (Deese, Lazarus, & Keenan, 1953; Taylor, 1958).

Complexity can be defined in terms of the number of components or steps which are essential to the performance of a task or solu-

[1] This study is based upon a dissertation presented to the Department of Psychology of the University of Chicago in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[2] The author wishes to express his grateful appreciation to Morris Stein, Donald Fiske, and Austin Riesen, who provided unfailing guidance and encouragement whenever it was needed.

Now at the State University of Iowa.

tion of a problem. Little research has been done on the interaction between complexity, defined in this way, and anxiety. Grice (1955) found that a high manifest anxiety group had a longer reaction time in a complex perceptual discrimination task, but the difference was accounted for by the lower intelligence scores of the high group. Using the Children's Manifest Anxiety (*CMA*) scale LA elementary school children have been found superior to HA children on difficult, complex discrimination learning tasks by Palermo, Castaneda, and McCandless (1956) and Castaneda, Palermo, and McCandless (1956). Initial differences in ability were not controlled in the latter studies; and none of these investigators systematically varied complexity as a function of the number of task components. As pointed out above, drive theory predicts more response competition and poorer performance for HA subjects as the number of components essential to a task are increased (Farber, 1955). The nondrive prediction would be of poorer performance by HA groups, *if* a given increase in the complexity or difficulty of a task is more frustrating or threatening to the HA subject and thus tends to reorient his behavior in an ego defensive direction (French, 1952; Saltz & Hoehn, 1957).

The purpose of this study is to determine the effects of anxiety upon performance on complex reasoning problems. Reasoning, in this instance, means the task of abstracting a relational principle from a model situation and applying it to a new situation. Tests of logical deduction and figure analogies are used. Complexity is varied by increasing the number of elements which have to be taken into account in formulating the principle of the problem. Within each test there are four subtests—each more complex than the one preceding. In order to isolate the effect of the complexity variable, mean scores on the complex subtests are adjusted relative to mean scores on simpler subtests by means of analysis of covariance. Three measures of manifest anxiety, provided by three different personality tests, are used. In addition, a measure of the "fear-of-failure" variant of anxiety has been included. The hypotheses are as follows.

Hypothesis 1. The performance of the high manifest anxiety group on complex problems shows a greater decrement from their performance on simple problems than is shown by the low manifest anxiety group.

Hypothesis 2. The performance of the high fear-of-failure group on complex problems shows a greater decrement from their performance on simple problems than is shown by low fear-of-failure group.

METHOD

Personality Variables

In an attempt to control test-specific variance, three types of personality tests[3] which have previously proven useful were used—a Self-Description Inventory and two kinds of projectives, a Sentence Completion test, and a Picture Interpretation test. Defining criteria were formulated to serve as the basis for rating the projectives and selecting the inventory items. The criteria were expressed in terms of the individual's typical reaction to frustration and attempt to cope with it. The definition of manifest anxiety includes any expression or implication of awareness of anxiety.

The defining criteria for manifest anxiety cover three areas: Frustration is perceived as a threat to the individual's *basic security*. Warning of the threat is given by an active state of emotional upset characterized by tension, anxiety, and apprehension. The primary aim of the individual is to protect himself from misery, attack, punishment, etc. The individual expects his *instrumental efforts* to resolve frustration to be confused, futile, thwarted, and ineffectual. His outlook is pessimistic, and his lack of hope for a constructive resolution may turn his efforts toward finding a way of escaping the frustrating situation. In general, *other persons* are mistrusted. The individual feels misunderstood and without a reliable person to whom he can turn for aid or reassurance. Authorities often reject or punish in spite of the individual's good intentions. Peers are expected to be troublesome or indifferent.

A modification of the Michigan Sentence Completion Test (Kelly & Fiske, 1951) was used as one instrument. Over half the original items were dropped, and first person references were changed to the third person or to an impersonal phraseology. A few items were added which referred to success, failure, and goals. The instructions were to complete all of the sentences, to write down the first thing that came to mind, and to work quickly. The time limit was 15 minutes.

The Picture Interpretation test was based upon 15 pictures clipped from old magazines. The pictures depicted one or more persons in ambiguous circumstances which could be interpreted as frustrating. The

[3] The instruments used in this study have been deposited with the American Documentation Institute. Order Document No. 6778 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $2.00 for microfilm or $3.75 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

subjects were told not to describe the picture, but rather to give a brief, imaginative interpretation of it by suggesting the thoughts, wishes, fears, intentions, beliefs, and feelings of the people which would be underlying reasons for the situation being the way it was. Each picture in turn was held in front of a group of subjects, and 70 seconds were allowed for writing the interpretation.

The following score categories were used for both projectives:

Score 2: if the themes and self-descriptive statements correspond to the criteria (in at least two of the main areas) so frequently and definitely that the variable (manifest anxiety) can be considered one of the most important characteristics of the individual's personality.

Score 1: if there is some evidence that the variable is very important in the personality under consideration, but for some reason the evidence is not decisive. (Note—this score is not to be used excessively. Unless there are hints that the variable might be very important, the "0" score should be used.)

Score 0: if the evidence indicates that the variable is not one of the primary characteristics of the individual's personality.

Indicate no confidence: in addition to one of the other scores when it is felt that there is so little evidence that the rating is apt to be invalid.

Each projective was rated by the author and one of two other judges. The ratings of the Sentence Completion and of the Picture Interpretation tests were made independently of each other in all cases. The judges had copies of the defining criteria for manifest anxiety as presented above. Each judge made several practice ratings and discussed them with the author until it was clear that the judge understood the procedure and the criteria for the variables. All raters were advanced graduate students in clinical psychology with internship experience.[4] The author divided each protocol into fourths (in the case of the Picture Interpretation test) or fifths (in the case of the Sentence Completion test). Each section of a test was rated for all subjects before going on to another section. With knowledge of his ratings of the separate sections, the author reread each protocol and made an overall rating on manifest anxiety. To save time for the other two judges, the protocols were divided only into two sections. The ratings of the two sections for a given projective were mechanically combined for a given judge. When two ratings were combined, either from a single judge or from two judges, the arithmetic mean was used except when the two ratings were a 1 and a 2. The combination of the latter two ratings was taken to be 2, since the score of 1 was to be used only when the judge felt that manifest anxiety might be very important for the person (that is, likely to be a 2), except that the evidence was not decisive. This convention helped to compensate for a marked regression toward the mean which occurred at the high end of the scale because of the very conservative set given to the judges.

Product-moment correlations were computed for interjudge agreement. On the Picture Interpretation test, the correlations between the author and each of the other judges were .68 and .62. On the Sentence Completion test the correlations were .38 and .41. These correlations were all significant beyond the .001 level except the last one, which was at the .01 level. The Picture Interpretation test correlations seem satisfactory given the complexity of the rating criteria. While the Sentence Completion test correlations are low, they are still worth consideration since two other measures of the same variable are available.

The ratings of the two judges on each protocol were combined. The Picture Interpretation and Sentence Completion tests each provided a separate manifest anxiety score. For testing the experimental hypothesis with respect to each of these scores, extreme groups were selected, consisting of all subjects with the score of 2 (high) or 0 (low). When "no confidence" was indicated by both judges, the subject was eliminated as far as that score was concerned.

The inventory was made up of 89 items to which a subject answered "true," "false," or "undecided," depending upon how he thought the item described himself. He was asked not to use the "undecided" alternative more than 10% of the time. Some of the items for the inventory were newly written; some were from the *MA* scale and some were need Achievement items. Twenty-four items were selected which seemed to fit the defining criteria for manifest anxiety. Eleven of these items also involved need for achievement, as well as anxiety, in a kind of fear-of-failure syndrome. Two provisional scales were scored consisting of the 11 fear-of-failure items and the 13 regular manifest anxiety items for all 200 subjects tested. Item analyses were done in order to determine which items discriminated between the high and low quartiles on each scale. It was found that quite a few items discriminated at the .001 level of confidence or beyond; therefore, this p level was selected as the cut-off point for selecting items. Using this stringent criterion for item selection produces scales which have high internal consistency. There were 26 items in the final Manifest Anxiety scale and 13 in the final Fear-of-Failure scale. The two scales are related; the three items which met the selection criterion on both scales had to be classified according to content.

Twenty-one of the manifest anxiety items were from Taylor's original scale. The five new items were either highly similar to the originals or extended the referent of the anxiety into the interpersonal domain, e.g., "People often disappoint me." Not included were items which refer to psychosomatic symptoms or to sleep difficulties, restlessness, and excitability without a connotation of manifest anxiety. Although the scale forms a self-consistent set, it is of interest to compare it with a factor analysis of the *MA* scale done by O'Connor, Lorr, and Stafford (1956). Most of the items used here were loaded on their Factor A, anxiety and insecurity, or Factor D, inferiority. Several of the items were loaded on the two tension factors, C and E; however, all of the "tension" items included had an explicit connotation of worry or anxiety. None of the items were pure on Factor B, autonomic instability. The items in

[4] Robert Yufit and Nelson Jones very kindly lent their time and skill to rating the projective data.

the Fear-of-Failure scale do not appear in the *MA* scale, but several of them closely resemble items in the Test Anxiety scale of Mandler and Sarason (1952).

For testing the experimental hypotheses with respect to the inventory scores, extreme groups were selected on the basis of cutting points established at salient discontinuities about one standard deviation above and below the mean on each inventory score. Subjects who marked "undecided" or who failed to respond on 20 or more items were rejected.

The intercorrelations among the four personality measures are shown in Table 1. A rank order coefficient was deemed appropriate because the projective scores included only a few gross intervals. Kendall's tau, which could readily be corrected for the many tied ranks that occurred, was used. The correlations were computed for each school-sex sample separately, and all subjects providing usable data on all personality measures were included. The inventory manifest anxiety score was found to be significantly related to the sentence completion manifest anxiety score and to the inventory fear-of-failure score in four of the five sample groups. The picture interpretation manifest anxiety score was significantly correlated with the sentence completion manifest anxiety score in only two of the samples. No other correlations were significant. Positive trends seemed apparent in the other relationships except for that between picture interpretation manifest anxiety and inventory fear-of-failure. There were no systematic differences in the correlations which could be attributed to school or sex.

TABLE 1

RANK CORRELATIONS (tau) AMONG PERSONALITY MEASURES IN FIVE SAMPLES

Personality Measure	Manifest Anxiety Measure		
	Sentence completion	Picture interpretation	Inventory
Picture Interpretation Manifest Anxiety Sample: (*N*)			
AM (22)	−.14		
BM (24)	.02		
CM (28)	.41**		
AF (24)	.09		
BF (33)	.33*		
Inventory Manifest Anxiety Sample:			
AM	.14	.03	
BM	.50**	.17	
CM	.41**	.24	
AF	.39*	−.08	
BF	.34*	.21	
Inventory Fear-of-Failure Sample:			
AM	.29	−.24	.52**
BM	.07	.11	.34*
CM	.10	−.02	.44**
AF	.27	−.02	.55**
BF	.11	.04	.23

* Significant beyond the .05 level.
** Significant beyond the .01 level.

In all cases, the relatively low magnitudes of the correlations indicated that the scores could not be substituted for one another. Therefore, a separate test of the hypothesis was made for each score. In addition, the hypothesis concerning manifest anxiety was tested with extreme groups selected on the basis of a combination of scores provided by the three instruments. Low groups of sufficient size were obtained by selecting those subjects who fell in the low group on all three instruments. This stringent criterion did not provide groups of sufficient size at the high end of the scale. Therefore, any subject who fell in the high group on two of the personality instruments was selected for the combined group. This procedure makes sense if one assumes that the instruments were rather insensitive and missed some high subjects.[5]

Reasoning Tasks

The reasoning tasks were tests of figure analogies and logical deduction, each of which consisted of 60 multiple-choice problems (cf. Thurstone, 1938). In either case, each problem consists of a set of logically interrelated components, the model, and a set of response alternatives, one and only one of which is logically consistent with the model set. Both tests are divided into four 15-problem subtests representing four levels of increasing complexity. Subtest I (simplest) was given first, and the level of complexity of the succeeding subtests was regularly increased so that Subtest IV (most complex) was given last. Several practice problems served as an introduction to Subtest I on each type of test. The four subtests of a given type are parallel except for complexity. Although a wide variety of elements and inference structure are used, each type of inference is represented by similar items in all four subtests. At all four levels, some of the problems are made more difficult by the presence of irrelevant information and nonsense terms.

In the Figure Analogies test, the model set consists of a pair of figures (1 and 2) which differ from each other on one or more figural dimensions: size, shape, shading, number, and position. A third figure (X) is shown, and the subject is required to select from five alternative figures (A to E) the one which bears the

[5] The attempt was made to study need Achievement defined in a way similar to that used in defining manifest anxiety. The three need Achievement scores were found to be uncorrelated with one another and essentially uncorrelated with all anxiety variables. Furthermore, none of the need Achievement variables singly or in combination were found to have consistent relationships with performance on the complex reasoning tasks. The results for the need Achievement variable are reported in Van Buskirk (1959).

It had been hoped to define high and low fear-of-failure groups on the basis of extreme scores on both variables. However, this was impractical because of the small number of subjects who clearly fell into such patterns. Therefore, our only measure of the fear-of failure variant of anxiety is that provided by a special inventory scale.

same relationship to the Figure X as Figure 2 bears to Figure 1. The inference required may be summarized as: 1 is to 2 as X is to ?. In the subtest at the simplest level (Level I), the two model figures differ in only one significant respect. In each succeeding subtest, the number of differences between the model figures is increased by one. In Subtest IV, four differences within the model pair of figures have to be taken into account.

In the logical deduction Subtest I, there are two propositions in the model set and three alternative conclusions. The complexity of each succeeding subtest is increased by the addition of one more proposition to the model set at each level, and by the addition of one more conclusion at Levels III and IV. In order to solve a problem in Subtest IV, the subject has to deduce a conclusion from a chain of five propositions and to identify the conclusion among five response alternatives. The complexity of the problems in Subtest IV is so great that they are extremely difficult.

Each subtest is timed separately with the time limit increased at each level. The times are short enough to prevent most subjects from completing any subtest. In order to reduce the effects of speed the scores have been corrected for guessing by subtracting the percentage of wrong answers from each subject's score which he would have gotten correct by chance. Each subtest provides a separate corrected score.

Subjects and Procedure

The subjects were undergraduate students from three different universities. Most of the students participated in this experiment in connection with some psychology course. As a result, they received formal or informal credit in that course for participation. Two-thirds of the students at School C responded to a notice in the dormitory offering $1.00 to participants in a psychological experiment. All prospective participants were told that five monetary prizes would be awarded for the five highest total scores on the two reasoning tests. First prize was $10.00, the second $5.00, the third $3.00, the fourth $2.00, and the fifth $1.00. The students were encouraged to participate but were essentially volunteers. At Schools A and B, both males (M) and females (F) served as subjects. At School C, only males participated in sufficient numbers to be included in the study. They ranged in age from 17–25. A few subjects who had ages of 27 or greater were excluded, as were those who failed to follow instructions or did not complete the testing.

Combining the characteristics of school and sex gave five sample groups: AM, AF, BM, BF, and CM. These five groups differed rather widely among themselves on many of the scores used in this study; therefore, the data have been analyzed for each of the five groups separately.

The tests were administered in group sessions. At School A, the testing was accomplished in a single 2-hour session in which the tests were given in the following order: Picture Interpretations, Sentence Completion, Inventory, Figure Analogies, Logical Deduction. At School B, students were tested in two successive regular psychology class sessions. In the first session, they were given Picture Interpretation, Sentence Completion, and Figure Analogies tests. In the second session, they took the Inventory and Logical Deduction tests. At School C, the paid subjects were tested in a single session and the others in two classes.

RESULTS

The hypotheses were tested by determining differences in performance in high and low groups on each of the personality variables. By use of the analysis of covariance technique, the means of the complex subtest were adjusted so that the variance associated with initial ability was partialed out; thus the proportion of the variance due to complexity was maximized. In the Figure Analogies test, Subtest I, for which the mean in all groups was considerably higher than the means of other subtests, provided a good measure of ability on simple items. Therefore, the Subtest I score was used as the control variable in the analysis of covariance. The means for Subtests II and III differed only slightly; for economy's sake, Subtest II means were not analyzed. In the Logical Deduction test, Subtest I, there was a large "warm-up" factor. The Subtest II mean was at least as great as the Subtest I mean in each group. It seemed advisable to reduce the proportion of variance due to warm-up in the measure of ability on simple items. Therefore, the sum of Subtests I and II was used as the control variable for the analysis of covariance of the logical deduction scores. The correlations between the two types of reasoning test were generally positive but low, so the data were treated separately for the Figure Analogies and Logical Deduction tests. Differences associated with school or sex were controlled by computing statistical tests of significance for each school-sex sample separately and combining the resulting p values for an overall test of significance, using the chi square technique suggested by Jones and Fiske (1953).

The first hypothesis was that the high manifest anxiety group would show a greater decrement on complex problems than the low manifest anxiety group. In Table 2 are shown the differences between the adjusted complex reasoning test means of the low and high groups selected on the basis of the combination of the three manifest anxiety scores. The differences were computed by subtracting the high anxiety group's mean from the low group's mean so that a positive difference would be in accord

with the hypothesis. One-tailed tests of significance were made. The differences on the Figure Analogies subtests were clearly within the range of chance variation. On the Logical Deduction subtests, the differences were consistently in the negative direction and yielded overall p values of .99 for Subtest III and .95 for Subtest IV. These results tended to contradict the hypothesis. In fact, the differences in the unexpected direction were large enough to demand that their statistical significance be tested by a two-tailed test. Within the theoretical context of this study, the possibility that the various samples showed significant differences in different directions was of no relevance. The alternative case of considerable interest was the proposition that all five sample groups showed differences contrary to our original hypothesis. In order to maintain consistency of direction, a one-tailed test of the difference in the unexpected direction was obtained for the five sample groups. These p values were then pooled, and the resulting value was doubled.

TABLE 2

ADJUSTED DIFFERENCES BETWEEN REASONING SUBTEST MEANS OF LOW AND HIGH GROUPS ON THE COMBINED MANIFEST ANXIETY CLASSIFICATION

($N = 52$)

Sample (school and sex)	Logical Deduction				Figure Analogies			
	Subtest III		Subtest IV		Subtest III		Subtest IV	
	Difference	p	Difference	p	Difference	p	Difference	p
AM	−1.3	.99	−0.8	.84	−1.4	.94	−1.6	.95
BM	−1.0	.89	−1.3	.94	+0.2	.44	+0.7	.34
CM	+0.1	.46	−0.3	.63	+0.5	.31	−0.4	.63
AF	−1.7	.97	−0.1	.52	−0.5	.78	+1.2	.12
BF	−2.7	.99	−2.0	.98	−0.2	.51	−0.5	.66
Combined p		.99[a]		.95[b]		.81		.60

[a] A two-tailed test of the difference in the direction opposite to hypothesis is significant, $p < .01$.

[b] A two-tailed test of the difference in the direction opposite to hypothesis is significant, $p < .05$.

The two-tailed test showed that the high manifest anxiety group was superior on both complex subtests of logical deduction. The differences were more extreme on Subtest III for three of the sample groups and, in fact, reached the .01 level overall. On Subtest III, the variances in Samples AF and BF approached heterogeneity ($p < .10$), but it is doubtful if this accounted for the overall result. However, there was no clear demonstration that the results were truly more extreme on one subtest than on the other subtest of logical deduction.

In Table 3 are shown the median mean differences and combined p values (resulting from combining results for the five samples) for the comparison between manifest anxiety groups selected on the basis of each separate personality instrument. For each type of personality instrument, at least one of the comparisons on logical deduction was in the

TABLE 3

SUMMARY OF ADJUSTED REASONING MEAN DIFFERENCES BETWEEN LOW AND HIGH GROUPS SELECTED ON THE BASIS OF THREE SEPARATE MANIFEST ANXIETY MEASURES

Manifest Anxiety Measure	N	Logical Deduction		Figure Analogies	
		Subtest III	Subtest IV	Subtest III	Subtest IV
Picture Interpretation	60				
Median difference		−0.8	−0.1	−0.8	−0.3
Range differences		−1.9 to +0.3	−2.0 to +2.1	−1.6 to +0.3	−1.3 to +0.5
Combined p		.06	—	—	—
Sentence Completion[a]	49				
Median difference		−1.0	−1.0	+0.6	+0.8
Range differences		−1.8 to +1.4	−1.1 to +0.2	−0.8 to +1.3	−1.2 to +1.9
Combined p		.05	—	—	—
Inventory	53				
Median difference		−0.1	−1.2	−0.1	−0.2
Range differences		−2.2 to 0.0	−2.5 to −0.1	−2.5 to +1.4	−1.1 to +0.4
Combined p		.05	.05	—	—

Note.—All significant differences are in the direction opposite to hypothesis.

[a] Based on four samples only; sample AM had no "highs" on this measure.

TABLE 4

ADJUSTED DIFFERENCES BETWEEN REASONING SUBTEST MEANS OF LOW AND HIGH GROUPS ON THE INVENTORY FEAR-OF-FAILURE SCALE

($N = 41$)

Sample (school and sex)	Logical Deduction				Figure Analogies			
	Subtest III		Subtest IV		Subtest III		Subtest IV	
	Difference	p	Difference	p	Difference	p	Difference	p
AM	−1.5	.95	−2.0	.97	−0.9	.81	+0.5	.31
BM	−0.1	.52	−1.8	.93	−0.9	.72	+0.6	.34
CM	−1.4	.93	+1.7	.07	+1.1	.16	+1.9	.08
AF	−1.0	.75	−1.2	.76	+0.6	.34	+1.0	.26
BF	+0.2	.43	+0.4	.39	+0.8	.22	+2.3	.04
Combined p		.95		.63		.44		.05

direction opposite to prediction at around the .05 level. The differences tended to be more extreme on Subtest III for the groups selected on the basis of the projective instruments. In the case of the figure analogies comparisons, the differences were all nonsignificant. The only new feature was that the difference between the groups selected by the sentence completion score showed a positive trend.

The differences between the adjusted complex reasoning test means of the low and high groups on the inventory fear-of-failure score are shown in Table 4. Again the high group's mean was subtracted from the low group's mean. We pointed out earlier that the fear-of-failure and manifest anxiety scores on the inventory are rather closely related. Fifty-six percent of the subjects involved in the comparisons on fear-of-failure were also involved in the comparisons on inventory manifest anxiety. Therefore, these results may not be considered independent of the findings on manifest anxiety. On the Logical Deduction test, the high fear-of-failure group tended to be higher on Subtest III, contrary to the original hypothesis. However, the two-tailed test of the difference in the unexpected direction was not significant. On Subtest IV of the Figure Analogies test, the high fear-of-failure group is inferior in all samples, and the combined p value is .05.

DISCUSSION

For any of three methods of measuring manifest anxiety, the high group was superior to the low group on some complex, verbal, logical deduction problems. Neither group was superior on complex figure analogies problems. Although the results vaguely suggested that the differences on logical deduction would disappear if the task were complex enough, the superiority of the high group seems more likely to be related to the increased complexity of the problems than to anything else. These results show that an increased number of elements in a reasoning task does not interact with the higher drive of the high manifest anxiety subject to produce greater interference among intratask (or extratask) responses. The extensive evidence (cf. Spence, 1958) that manifest anxiety is a measure of drive is not necessarily challenged. We need a better theory of the way in which drive, habit, and response variables interact in a multiresponse task when no one response has great initial strength. Perhaps a crucial factor is that reasoning tasks call for each correct response to be used only once instead of being repeated in a certain order as in the learning tasks previously used. At any rate the significance of the nature of the task for predicting the effects of motivation is emphasized.

Another theory that merits consideration is that differential motivational states in HA and LA subjects orient them in different directions toward the experimental tasks. It is clearly not the case that an increase in the complexity or difficulty of a reasoning task is unusually discouraging or disorganizing to the high manifest anxiety subject. On the contrary, it would seem that the HA subject is trying harder. There is evidence that the HA subject is more responsive to experimental reward (Taffel, 1955) but more sensitive to an ego involving threat. In two experiments, Sarason (1956, 1957) found the HA group to be superior to the LA group only when neutral instructions were used. The task was a very difficult list of nonsense syllables on which no group achieved a perfect recitation. When ego involvement was introduced through a threat that the results on the experiment would affect future evaluations of the subjects intelligence, the LA group was significantly superior to the HA group. The persistence with which an individual works toward a goal is a joint function of his need for the goal and his expectation

of obtaining it under given conditions (French, 1952). When the experimental set is neutral or emphasizes service to the experimenter for a monetary incentive, as in the present experiment, the HA subject would have a greater commitment to performing well because of his generalized need for reassurance and approval. When the threat of a negative evaluation contingent upon successful performance is introduced, both the need and expectation of success of the LA subject should change in the direction of greater commitment to successful performance. However, the HA subject's great self-doubt, or expectation of frustration and punishment, should interfere with an objective, need persistent orientation. The performance of an HA group under ego involving instructions should show the characteristic decrements shown by all groups after an arbitrary report of failure (Lazarus, Deese, & Osler, 1952; Taylor, 1958). The more inferior performance of the high fear-of-failure group in the present experiment is consistent with this theory. The high fear-of-failure subject describes himself through the inventory as having a much greater expectation of failure—particularly in any testing situation.

A final consideration is whether the anxiety variable is interacting with task subject matter. In a minority of the figure analogies problems, there were wrong response alternatives which looked more similar to the model set, from which the relationship is to be abstracted, than did the correct response alternative. The presence of such similarities introduces an element of prepotent error associations which have repeatedly posed difficulties for high manifest anxiety subjects. An *ad hoc* classification of the two types of items showed that they occurred about equally often at each level of complexity. A separate scoring of each type of item revealed no consistent tendency for the HA group to make more errors on the type with interfering wrong responses. A related finding has been reported by Himmelweit (1945). He compared the Raven Progressive Matrices scores and vocabulary scores of neurotic subjects. The Raven Progressive Matrices is a test of analogical thinking in which a concrete set usually leads systematically to the wrong answer. Himmelweit found a significant tendency for anxious and depressed patients to have higher vocabulary scores and for hysterical patients (who deny anxiety) to have higher matrices scores. Palermo et al. (1956) and Castaneda et al. (1956) found differences favoring the LA group on complex, nonverbal, discrimination learning tasks. The question remains as to whether the general inferiority of anxious subjects on nonverbal tasks is a function of differential response to subject matter (perhaps threat evoked by novelty) or of the differential effects of intratask interference. A study needs to be done utilizing two types of nonverbal task, one with competing false response alternatives and the other with noncompeting alternatives.

SUMMARY

This study attempted to determine the effects of anxiety upon performance on complex reasoning problems. Manifest anxiety was measured with three instruments: picture interpretation, sentence completion, and an inventory. Two reasoning tasks, logical deduction and figure analogies, were each divided into four separate subtests of increasing degrees of complexity defined in terms of the number of components involved in the problem. It was found, contrary to hypothesis, that the high manifest anxiety groups performed at a superior level on logical deduction complex problems than did the low manifest anxiety groups. The differences were found for extreme groups selected on the basis of each separate personality instrument as well as on a combination of all three. There were no significant differences on the figure analogies problems for the manifest anxiety groups. The high group on an inventory "fear-of-failure" score was significantly inferior on the most complex figure analogies subtest. The results do not support the hypothesis that the performance of the HA subject will show greater intratask interference, disorganization, or ego defensiveness when the complexity of a task is increased simply by increasing the number of components in the task. The results are consistent with the interpretation that the HA subject is more strongly committed to performing well as long as the experimental situation does not present a serious, ego involving threat.

REFERENCES

CASTANEDA, A., PALERMO, D., & MCCANDLESS, B. Complex learning and performance as a function of anxiety in children and task difficulty. *Child Develpm.*, 1956, **27**, 327–332.

CHILD, I. L. Personality. *Annu. Rev. Psychol.*, 1954, **5**, 149–170.

DEESE, J., LAZARUS, R. S., & KEENAN, J. Anxiety, anxiety reduction, and stress in learning. *J. exp. Psychol.*, 1953, **46**, 55–60.

FARBER, I. E. The role of motivation in verbal learning and performance. *Psychol. Bull.*, 1955, **52**, 311–327.

FARBER, I. E., & SPENCE, K. W. Complex learning and conditioning as a function of anxiety. *J. exp. Psychol.*, 1953, **45**, 120–125.

FRENCH, T. M., *The integration of behavior*. Vol. 1. Chicago: Univer. Chicago Press, 1952.

GRICE, G. R. Discrimination reaction time as a function of anxiety and intelligence. *J. abnorm. soc. Psychol.*, 1955, **50**, 71–74.

HANFMANN, EUGENIA. Psychological approaches to the study of anxiety. In P. H. Hoch & J. Zubin (Eds.), *Anxiety*. New York: Grune & Stratton, 1950. Pp. 51–69.

HIMMELWEIT, H. T. The intelligence-vocabulary ratio as a measure of temperament. *J. Pers.*, 1945, **14**, 93–105.

JONES, L. V., & FISKE, D. W. Models for testing the significance of combined results. *Psychol. Bull.*, 1953, **50**, 375–382.

KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. Michigan Press, 1951.

LAZARUS, R. S., DEESE, J., & OSLER, SONIA. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

MANDLER, G., & SARASON, S. B. A study of anxiety and learning. *J. abnorm. soc. Psychol.*, 1952, **47**, 166–173.

O'CONNOR, J. P., LORR, M., & STAFFORD, W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, **12**, 160–163.

PALERMO, D., CASTANEDA, A., & MCCANDLESS, B. The relationship of anxiety in children to performance in a complex learning task. *Child Develpm.*, 1956, **27**, 333–337.

RAMOND, C. Anxiety and task as determiners of verbal performance. *J. exp. Psychol.*, 1953, **46**, 120–124.

ROSENZWEIG, S. An outline of frustration theory. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 1. New York: Ronald, 1944. Pp. 379–388.

SALTZ, E., & HOEHN, A. J. A test of the Taylor-Spence theory of anxiety. *J. abnorm. soc. Psychol.*, 1957, **54**, 114–117.

SARASON, I. G. Effect of anxiety, motivational instructions and failure on serial learning. *J. exp. Psychol.*, 1956, **51**, 253–260.

SARASON, I. G. Effect of anxiety and two kinds of motivating instructions on verbal learning. *J. abnorm. soc. Psychol.*, 1957, **54**, 166–171.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

TAFFEL, C. Anxiety and the conditioning of verbal behavior. *J. abnorm. soc. Psychol.*, 1955, **51**, 496–501.

TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285–290.

TAYLOR, JANET A. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57**, 55–60.

TAYLOR, JANET A., & CHAPMAN, JEAN P. Paired-associate learning as related to anxiety. *Amer. J. Psychol.*, 1955, **68**, 671.

TAYLOR, JANET A., & SPENCE, K. W. The relationship of anxiety level to performance in serial learning. *J. exp. Psychol.*, 1952, **44**, 61–64.

THURSTONE, L. L. Primary mental abilities. *Psychometr. Monogr.*, 1938, No. 1.

VAN BUSKIRK, C. Performance on complex reasoning tasks as a function of manifest anxiety and, need achievement. Unpublished PhD dissertation, University of Chicago, 1959.

(Early publication received January 9, 1961)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 210-215

ANTI-SEMITISM, JUDGMENTAL PROCESSES, AND DISPLACEMENT OF HOSTILITY[1]

LEONARD BERKOWITZ
University of Wisconsin

In an earlier paper (Berkowitz, 1959) the writer reported that anti-Semitic college students were more likely to displace hostility onto a previously neutral bystander when angered than were their less prejudiced peers. These findings not only tend to support the scapegoat theory of prejudice; they also raise the problem of explaining the negative results obtained in other, similar investigations. Lindzey (1950), for example, exposed groups of subjects, either very high or very low in minority group prejudice, to experimentally-induced frustrations and found that the prejudiced subjects did not differ significantly from the less prejudiced group on his measures of hostility displacement.

Since the writer's investigation had also used extreme groups selected on the basis of a very similar measure of anti-Semitism, it seems likely that the variation in results can best be accounted for in terms of the indices of displacement. Berkowitz used questionnaire ratings of the subjects level of friendliness toward a given peer, the experimenter's confederate, while the other experimenters employed fantasy measures of aggression. This difference may be important because the ratings conceivably could have been more affected by judgmental processes than were the fantasy indices. Holmes and Berkowitz (1961) demonstrated that ratings of friendliness are subject to judgmental influences; conditions analogous to those producing judgmental contrast effects in psychophysical experiments gave rise to contrast-like effects in these friendliness ratings. However, granting the operation of judgmental processes, we also must assume that highly prejudiced people make different kinds of judgments when frustrated than do their less prejudiced fellows.

A recent paper has shown how these differences might operate (Berkowitz, 1960). As is well known, when a person makes a judgment he compares the to-be-evaluated stimulus with some appropriate standard. A contrast effect arises in this process when there is a relatively great psychological distance between the stimulus and the standard, i.e., the distance appears greater than it "really" is (as defined, for example, by an observer not affected by the standard), while assimilation, in which this distance is minimized, occurs when the true distance is relatively small. Psychological distance can be coordinated, among other things, to the number of discrete categories the observer can establish between the to-be-evaluated stimulus and the standard on relevant dimensions. Generally, the greater the number of discriminations the observer can make among stimuli objectively located between the critical stimulus and the standard on these dimensions, the greater the perceived distance between them. Persons who typically employ broad categories when evaluating stimulus objects should therefore be more likely to display judgmental assimilation than people who characteristically make finer discriminations.[2]

In applying this analysis to the study of anti-Semitism and displacement, the writer has suggested that stress serves to broaden the categories established by the highly prejudiced and to narrow those set up by the less prejudiced. As a result, the former are likely to exhibit assimilation. They do not clearly differentiate between the immediate frustrator and the objectively neutral person so that the negative feelings aroused by one can be generalized readily to the other. (This differentiation probably would take place, however, when the

[1] This study, conducted by the writer's class in Experimental Social Psychology, is part of a research program supported by Research Grant M1540(C2) from the National Institute of Health, United States Public Health Service.

[2] This statement can be tautological in terms of some operations. For example, "broad categories" can be taken to mean a relatively flat stimulus generalization gradient for stimuli in that class. The person responds the same way to all of the objects in the category because they have similar stimulus qualities for him. Similarly, "assimilation" also involves stimulus generalization. In this case the same judgmental responses are made to the objects because they have similar stimulus qualities.

stimulus objects are a "real" person, the experimenter, and a "nonreal" picture on a TAT card.) The less prejudiced, on the other hand, apparently respond to the stress of the frustration by differentiating the frustrator from the bystander more sharply, a tendency heightening the contrast between these two objects and, according to the results, they become somewhat friendlier to this bystander.

Two studies were used to support this line of thought. In the first of these, Brown (1953) showed that highly authoritarian persons (as defined by the California F Scale) exhibit greater problem solving rigidity than less authoritarian subjects, but only when placed in the stress of an ego involving situation. When this finding is translated into the present theoretical framework, it suggests that the authoritarian's categorizations generally become broader under at least mild stress. Once stress is created, he becomes less likely to differentiate into the various types of problems that he could discriminate among under conditions of low ego involvement. Apparently, as far as he is concerned, they all are in one category, and he responds in a similar manner to all of the problems.

The second study, conducted by Pauline Schoenberg under the present writer's supervision, sought to test this interpretation employing the format of a more typical investigation of judgmental processes. The subjects were asked to rate the "goodness" of 11 personality traits, with the first 7 traits, all clearly "good" or all clearly "bad," as defined by the prior judgments of 68 other subjects, providing the comparison standard, and the remaining 4 being relatively neutral. Other investigations employing these traits had shown that there generally is a contrast effect in the evaluation of the neutral traits; they are judged as relatively "bad" when preceded by the 7 good traits, and as relatively "good" when they follow the bad standard. In other words, most subjects seem to differentiate the neutral traits from the preceding traits so that by comparison with the latter standard they appear more different than they really are. Following this reasoning, we hypothesized that subjects high in F would be less likely to show a contrast effect in their evaluations of the neutral traits when placed in an ego involving condition than when given less ego involving instructions, supposedly because the stress of the former condition caused them to make poorer discriminations between the standard and neutral traits. The earlier displacement experiment also suggested that these results might be reversed in the case of less authoritarian subjects. In that study, it will be recalled, subjects who were low in prejudice became somewhat friendlier to the neutral peer when they were frustrated by the experimenter, as if they had sharpened their differentiation between these two. Extrapolating, we predicted that subjects low in F would be more likely to exhibit a contrast effect when ego involved than when not given ego involving instructions. The results of this second study generally were consistent with our expectations (Berkowitz, 1960).

However, there are difficulties in extending these findings to the relationship between anti-Semitism and displacement of hostility. The instructions given the subjects in the judgment experiment emphasized that the traits all referred to one person. Although the writer has argued against this possibility (Berkowitz, 1960), the obtained results conceivably could stem from differences in adherence to the "all-one-person" instruction. The present study was designed to resolve this problem. The subjects selected primarily in terms of their scores on the anti-Semitism scale are asked to evaluate the same sets of traits employed in the earlier Schoenberg-Berkowitz study, but without being told that the traits pertain to one particular person. Nevertheless, essentially the same results are expected: the highly anti-Semitic subjects should display grosser categorizations under the stressful situation than in the more secure, non-ego-involving situation, while the reverse should be true for the subjects low in anti-Semitism.

METHOD

Subjects

Sixty-five female and 33 male college students participated in the two sessions of this experiment. They were recruited from a pool of approximately 120 volunteers from an introductory psychology course who, in a preliminary session, had completed a 111-item, five-alternative, Likert questionnaire supposedly assessing their "attitudes toward everyday matters." Twenty-eight of these items were taken from the California F Scale, while 17 items came from the measure of anti-Semitism (A-S) developed by the same

investigators (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950). In both cases item scores varied from four points for a response indicating strong agreement with an item to zero for the response "Strongly disagree." The distribution of obtained scores on these measures was concentrated toward the low end of the possible range: 12 to 67, median = 44 for the F Scale, and 1 to 49, median = 14 for the A-S scale.

In the preliminary session the subjects also filled out a separate work and class schedule, giving their names and addresses. The writer coded these papers so that the 15 experimenters taking part in the study[3] would not know the subjects' attitude scores and randomly assigned each subject to one of the eight experimental conditions, maintaining controls only on the cell frequencies and the ratio of men to women in each condition. Each experimenter then contacted one of the six or seven subjects assigned to him for the second session, when he administered the instructions appropriate to the condition for which the subject was scheduled. Knowing only the F and A-S scores of the subjects who had participated in the second session, but not the nature of their judgmental responses, the writer then selected those subjects to be included in the data analysis. The highly anti-Semitic subjects were those falling in the upper third of the A-S distribution of the sample (a score of 17 or higher), and also above the median on the F Scale, while the less prejudiced subjects were those who came from approximately the lower third of the A-S distribution (a score of 12 or lower), and were below the median on the F Scale. As a result, there were 56 subjects divided evenly among the eight experimental conditions in a $2 \times 2 \times 2$ factorial design.

Experimental Procedure

In the second session the experimenter told each subject that the purpose of the study was to develop quantitative ways of evaluating personality descriptions, such as those commonly included in letters of reference. Supposedly the experimenter wanted to find out how "good" or how "bad" the adjectives were that were frequently used in these descriptions. To obtain this information, the experimenter would read a list of personality traits to the subject and the subject would indicate the "goodness" of each trait by calling off a score for each adjective. A score of 1 was to mean that the trait was a "very good" personal quality, while 13 meant the trait was "very bad."

Half of the subjects were placed in a mildly stressful, ego involving situation by informing them that their ratings would constitute a test, involving comparison with the evaluations made by experienced personnel directors in order to determine which subjects were good judges. (This supposedly was preliminary to isolating the personality characteristics of good judges.) A lower degree of ego involvement, and therefore, presumably of stress, was induced in the remaining subjects by telling them that their judgments were not a test of their personality, or of how good a judge they were. The experimenter said he only wanted to see how "people in general" felt about these traits.

Eleven traits then were read to the subject, with the subject giving a score after each adjective. The first seven traits provided the comparison standard. As reported earlier, 68 other undergraduates in a preliminary investigation had rated a number of adjectives on a 13-point scale. Half of the present subjects received as this standard seven traits that the judges had evaluated as "good," while the others heard seven adjectives the judges had evaluated as "bad." Immediately following the standard traits, and without setting them off as different in any way, the experimenter read the same four traits, which had previously been defined by the judges as "neutral,"[4] to all subjects. The traits and their scale values are given in Table 1. The standard traits were always given in the order listed in the table, but the order of presentation of the neutral traits was systematically varied according to a prearranged schedule.

RESULTS

The results of the Duncan multiple range test (Duncan, 1955) of the condition means for both the seven standard and four neutral traits are given in Table 2. A number of conclusions are readily apparent. First, there are no significant differences among the A-S level and stress level conditions in the mean judgments of the first seven traits within each of the standard conditions. Thus, all subjects getting either the "good" or "bad" comparison traits apparently received psychologically similar judgmental-standard stimuli. The question now is to what extent did the groups vary in their differentiation of the neutral traits from these traits?

Generally, this differentiation apparently was made, and there was a very strong contrast effect; when all subjects are combined, the neutral traits have a significantly poorer evaluation following the "good" comparison standard than after the "bad" traits ($F = 146.26$, $p = .001$). However, the lower half of Table 2 also shows that there are condition differences in the extent to which the neutral traits were discriminated from the standard stimuli. This is indicated, for example, in the

[3] Needless to say, the writer was pleasantly surprised that there were any significant findings, much less that there was the close conformity to the results obtained in the earlier Schoenberg-Berkowitz study.

[4] There was relatively good agreement among the 68 judges as to the ratings given to the standard traits, i.e., they have a relatively small Q value in the Thurstone sense. The "neutral" traits, on the other hand, have a relatively large Q value. Other studies carried out by the writer had indicated that ambiguous traits generally are more sensitive to judgment shifts.

judgmental differences between the "good" and "bad" standard groups within each of the four stress-level conditions. As expected, the subjects high in A-S exhibited a greater contrast effect under low stress than in the stronger stress condition. That is, only in the low stress treatment were the neutral stimuli evaluated significantly less favorably after the "good" than after the "bad" traits. Also as predicted, this pattern is reversed for the Low A-S subjects who display the significant contrast effect under high rather than low stress. In other words, the prejudiced subjects seem to make poorer discriminations between the standard and subsequent stimuli under mild stress than in the more neutral condition, while the reverse is true for the less prejudiced group.

These discrimination differences do not occur with all of the stimuli, however. As can be seen in the table, the evaluations of the neutral traits by subjects high in A-S are affected reliably only when these adjectives follow the "good" comparison standard. In the case of the less prejudiced subjects, on the other hand, there is a near-significant difference ($p = .07$) between the high and low stress conditions only for the neutral traits following the "bad" standard. (Since this difference was significant in the earlier Schoenberg-Berkowitz study, the finding is probably reliable.) Obviously, we cannot say at this point whether these results severely restrict the applicability of the present hypotheses. The fact that the High A-S subjects tended to exhibit relatively gross categorizations under stress only in the "good" conditions may be explained in many ways. For example, other meaning dimensions, such as the "potency" or "activity" (Osgood, Suci, & Tannenbaum, 1957) of the "bad" traits may have affected the extent to which the prejudiced individuals could categorize these adjectives together with the subsequent neutral adjectives; the two sets of traits may have been too far apart on these dimensions, at least as far as the High A-S subjects were concerned, for them to be placed in the same broad category. A similar account could be given of the behavior of the Low A-S subjects. Nevertheless, in both cases the present hypothesis seems generally to be supported: prejudiced subjects manifest a different kind of judgmental process under stress than do their less prejudiced peers.

TABLE 1

THE SCALE VALUES OF THE STANDARD AND NEUTRAL TRAITS

(1 = "very good"; 13 = "very bad")

"Good" Standard		"Bad" Standard	
Adjective	Scale Value	Adjective	Scale Value
1. Adaptable	1.57	1. Aloof	10.09
2. Ambitious	1.69	2. Downhearted	10.45
3. Cheerful	1.57	3. Belligerent	11.14
4. Intelligent	1.36	4. Glum	11.09
5. Confident	2.19	5. Irritable	11.59
6. Diligent	2.39	6. Dull	11.75
7. Warm	2.63	7. Obstructive	10.88

"Neutral" Traits

Adjective	Scale Value
8. Quiet	6.70
9. Reserved	6.12
10. Shrewd	6.05
11. Nonchalant	6.79

TABLE 2

MEAN JUDGMENTS IN THE EXPERIMENTAL CONDITIONS

($N = 7$ in each condition)

A-S Level:	High		Low		High		Low	
Stress Level:	High		Low		High		Low	
Standard:	Good	Bad	Good	Bad	Good	Bad	Good	Bad
Mean judgments of standard traits	2.66_a	9.87_e	3.31_a	10.94_e	2.57_a	10.01_e	2.56_a	10.10_e
Mean judgments of neutral traits	5.61_{be}	5.23_b	6.97_d	5.29_b	7.24_d	5.44_{be}	7.06_d	6.49_{cd}

Note.—Cells containing the same subscript are not significantly different from each other (at the .05 level) by Duncan multiple range test. All of the cells in this table were included in the one analysis. Note that even if the relatively slight mean anchorage differences were assumed to be reliably different, they could not in themselves consistently account for the differences in judgments of the neutral traits.

DISCUSSION

The present argument obviously faces problems in addition to the difficulties inherent in interpreting the judgmental results as stemming from differences in discrimination tendencies. In order to generalize the findings to the earlier study of hostility displacement it is necessary to equate the ego involving treatment in this experiment to the frustration produced by the experimenter in the other investigation. A recent paper by Easterbrook (1959) provides support for this perhaps unattractive assumption—and also presents other difficulties.

Easterbrook (1959) reviews a number of studies demonstrating "that emotional arousal acts consistently to reduce the range of cues that an organism uses" (p. 183). This reduction in "the range of cue utilization" is essentially identical to the present conception of "category broadening." The range is said to have shrunk when the organism makes relatively little use of peripheral (partially relevant) cues. Failure to attend to peripheral cues also implies that the organism responds primarily to the gross aspects of the stimuli and not make fine discriminations among them. In the present terminology, the stimuli are more likely to be grouped together in the same category. Not making these sharp differentiations (not responding to the fine details that differentiate the stimuli from each other), the organism acts as if the stimuli were largely equivalent.

Easterbrook (1959) points out that reduction in the range of cue utilization apparently takes place following a great variety of instigating conditions, all of which may be regarded as emotion arousing. These conditions include food deprivation, threats to the self or to the body, and what is particularly important for this paper, the offering of a prize for effective performance, which seems essentially similar to the ego involving instruction of the present study. In both cases performance was made much more important to the individual. Presumably, emotion might also have been aroused in these conditions, thus, equating these instigations with the frustration created by the experimenter in the hostility experiment, at least as far as the discrimination process is concerned.

Applying this reasoning to the present study, we also would have to contend that the apparently gross discriminations made by the High F subjects in the stress condition resulted from emotional arousal. Their behavior in the judgmental task then becomes a special instance of a more general phenomenon: a reduction in the range of cue utilization (i.e., category broadening) under emotion-producing conditions resulting in less adequate discrimination among the stimulus objects confronting them. As a consequence, responses (such as hostility) instigated by one of these objects could be generalized readily to other objects.

It is the behavior of the subjects low in F that is inexplicable from this point of view. One possibility is that the phenomenon does not occur in such people under stress. It may also be that the Low F people were not made emotional by the present conditions. However, such an explanation cannot account for the presumably *sharper* differentiations made by these subjects both under stress and following the experimental frustration.

SUMMARY AND CONCLUSIONS

The present study was designed to test a theoretical argument used to explain the different results in various investigations of the scapegoat theory of prejudice. Berkowitz (1959) had obtained evidence supporting the scapegoat theory, while other experimenters (e.g., Lindzey, 1950) have obtained essentially negative results. Briefly, it is suggested that the index of hostility displacement employed by Berkowitz was more affected by judgmental processes than the indices of fantasy aggression used by the other investigators, and that highly prejudiced people make different kinds of judgments under stress than do people lower in prejudice. According to this judgmental analysis, the conceptual categories established by those high in prejudice tend to broaden under stress, i.e., they make poorer discriminations, so that both the frustrator and a neutral bystander are more likely to be included in the same category and the hostility aroused by the former can readily be generalized to the latter. Less prejudiced people, on the other hand, apparently react to stress with a tendency to make finer differentiations. As a result, there was a heightened contrast between the frustrator and the bystander for these people and

their friendship for the bystander tended to increase.

To test these hypotheses, extreme groups on the Levinson anti-Semitism scale were placed either in a mildly stressful, ego involving condition or in a less stressful, non-ego-involving condition, and were asked to evaluate the general "goodness" of 11 personality traits. The first 7 (comparison standard) traits objectively differed from the last 4 traits on this "goodness" dimension. As was found in a preliminary investigation employing the same procedure (Berkowitz, 1960), the results generally are in accord with the hypotheses; the highly anti-Semitic subjects seemed to make grosser discriminations under stress than in the more neutral situation, while the reverse was true for the less prejudiced subjects. The findings suggest, furthermore, that these differences stem from differences in perceptual discriminations rather than from differences in degree of consistency, i.e., differences in a tendency to give the same evaluative responses. All subjects apparently differentiated the last 4 from the first 7 traits; their evaluations of these two sets of traits were different in the range of numbers used. Instead of differing in the consistency of their responses, the subjects seemed to vary in the magnitude of the perceived difference between the comparison and subsequent stimuli.

If this judgmental analysis is correct, it raises the problem of whether the questionnaire ratings employed in the writer's earlier study of displacement really assessed the strength of the subjects hostility drive. Undoubtedly, aggressive behavior and judgmental evaluations represent different response classes. Nevertheless, there probably is a correlation between these two sets of responses, at least at the extremes. Feelings of friendliness towards another not only result in more favorable evaluations of him than do feelings of neutrality or unfriendliness (Berkowitz, 1957), but the favorable evaluations may also weaken if not prevent the occurrence of aggressive acts. Moreover, with overt aggression generally frowned upon in civilized society, the expression of unfavorable evaluations frequently is the only relatively permissible way that aggressive inclinations may be displayed. It certainly should not be concluded that unfavorable evaluations have nothing to do with aggression.

REFERENCES

ADORNO, T., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D., & SANFORD, R. *The authoritarian personality.* New York: Harper, 1950.

BERKOWITZ, L. Liking for the group and the perceived merit of the group's behavior. *J. abnorm. soc. Psychol.*, 1957, **54**, 353–357.

BERKOWITZ, L. Anti-semitism and the displacement of aggression. *J. abnorm. soc. Psychol.*, 1959, **59**, 182–187.

BERKOWITZ, L. Judgmental processes in personality functioning. *Psychol. Rev.*, 1960, **67**, 130–142.

BROWN, R. A determinant of the relationship between rigidity and authoritarianism. *J. abnorm. soc. Psychol.*, 1953, **48**, 469–476.

DUNCAN, D. Multiple range and multiple F tests. *Biometrics*, 1955, **11**, 1–45.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66**, 183–201.

HOLMES, D. S., & BERKOWITZ, L. Some contrast effects in social perception. *J. abnorm. soc. Psychol.*, 1961, **62**, 150–152.

LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.

OSGOOD, C., SUCI, G., & TANNENBAUM, P. *The measurement of meaning.* Urbana, Ill.: Univer. Illinois Press, 1957.

(Received January 11, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 216-220

THE EFFECT OF AMOUNT OF POSSESSION OF THE TRAIT OF HOSTILITY ON ACCURACY OF PERCEPTION OF HOSTILITY IN OTHERS

BERNARD I. MURSTEIN[1]

Interfaith Counseling Center, Portland, Oregon

One of the basic assumptions of most personality theories is that a person's perception of others is influenced by his own personality characteristics. Further, if the possessed characteristic is negatively valued, the person's ability to assess the presence of the same trait in others is believed to be impaired. Insight into one's own personality is thought to mitigate such distortion. Indeed, one of the goals of didactic therapy is to help the therapist-in-training to gain insight into his own actions and thus avoid projecting his own motives onto his patients.

Despite the very general acceptance of these beliefs, they lack adequate experimental validation. The present study was therefore designed to investigate how possession of a trait of low social value affects the objective evaluation of the strength of the trait in others. A second objective was to determine whether insight into one's standing on a trait increases the objectivity with which a person assesses it in others.

The trait dimension of "hostility-friendliness" was selected as the negative trait to be studied, partly for reasons extraneous to this study (Murstein, 1956). There are, however, good reasons for choosing this dimension when working with fraternity members who comprised the subjects in this investigation. First, requests for judgments of this trait do not arouse undue opposition from the subjects because of reluctance to say anything unfavorable about the other members of their group.[2] Second, the trait can easily be ranked by subjects who are not professional psychologists.

At first glance the assumption that the persons ranked as least friendly, as in the present study, are also perceived as most hostile may seem questionable. May not a person be aloof and unfriendly and yet not hostile? But the manner in which "least friendly" was defined by the investigator made it clear to the subjects that these persons were to be evaluated as hostile. Subjects were told that "most friendly" meant "that person who is the most cooperative, easiest to get along with, and *least hostile* in the fraternity." Moreover, the magnitude of the average reliability for the group judgments of the 25 groups (.86) indicated that the rankers probably employed a unidimensional scale.

In studying possession of the trait of hostility-friendliness, the pooled-rank method was used. A prime advantage of this method is that each person uses the same distribution to evaluate the group, avoiding "central tendencies," "leniency errors," and "halo effects" that reduce discrimination among the members. In the pooled-rank method, if the "forcing" distorts the data the reliability of judgments by the group should be low. A fairly high reliability indicates that little distortion has occurred. A second advantage is that the method is easily explainable to the subject, requires little time to administer, and thus permits the use of large groups. A third advantage is that the rankings may be easily converted to normalized scores. The justification for this conversion rests on the assumption that most personality traits are normally distributed, a view championed by Guilford (1954) and Calvin and Holtzman (1953), among others. Although groups of fraternity students cannot be assumed to represent the population

[1] I should like to express my thanks to the following persons for their aid in the collection of data: Martin Dworkin, Allen I. Laskin, Joan Laskin, George Lassen, Joseph Marin, Nelly K. Murstein, and Roberto Roth. Maureen Bendick helped with the computations. Martha Pareis aided greatly in editorial matters relating to the preparation of the manuscript. Thanks are due to Charles A. Moss for kindly giving of his time in drawing the figure.

[2] In a previous "pilot study," the trait of "hostility" had been used. The use of this trait, however, met with considerable resistance on the part of the subjects, who were reluctant to term any of their fraternity brothers as hostile. Consequently, the trait deemed at the other end of the continuum, "friendliness," was employed, with the assumption that those subjects who were the least friendly were the most hostile.

at large for which this assumption is most justifiable, it seems unlikely that their distribution for the trait of hostility-friendliness differs sufficiently from that of a normal population to offset the statistical gain in working with normalized ranks (interval data) rather than with the ranks untransformed.

The purpose of the ranking was to select four extreme groups differing both in their possession of hostility and in the degree of insight into their possession or lack of possession of hostility. The hypotheses advanced were:

1. "Hostile" persons are less accurate in their perception of the trait of "hostility" in others than are "friendly" persons.
2. "Insightful" persons are more accurate in their perception of "hostility" in others than are "noninsightful" persons.

METHOD

The subjects were drawn from 23 fraternities and 2 men's dormitories at the University of Texas. Prior to a regularly scheduled "house" meeting, sheets of paper were handed out to all members present who had been members of the group for at least 3 months. The familiarity of each subject with all others in his group was thus assured. The names of all the members present for the meeting were written down by each member on his sheet. The following instructions appeared at the top of the subject's sheet:

> Put a 1 in the small column to the right of the name of the person whom you consider to be the most friendly of those taking the test with you. By "most friendly" is meant that person who is the most cooperative, easiest to get along with, and least hostile in the fraternity. Put a number 2 next to the second most friendly person, and so on, until you have a different number for each person currently taking the test. Remember to rate yourself and to make sure that your name is also at the top of the page, Also, do not rate any member who is not taking the test with you.

The ranks were converted to normalized scores, using Hull's table (Guilford, 1954, p. 182).

From the 536 students participating in the study, 80 subjects were to be selected who, on the basis of the group and self-judgments, could be assigned to one of four groups: hostile-insightful, hostile-noninsightful, friendly-insightful, friendly-noninsightful. To ensure that the subjects chosen for the four experimental groups were either very hostile or very friendly, criterion scores based on consideration of the standard error of measurement of each of the group judgments were established, such that the 20 most extreme cases exceeding these scores for each of the four experimental groups were selected. Split-half reliability coefficients were obtained by dividing members of each fraternity and dormitory group randomly into two halves and correlating the means of the ranking by each half for each person in the group, then applying the Spearman-Brown correction. The scores for each fraternity and dormitory are henceforth referred to as G scores.

The reliability of the self-scores (S scores) was obtained in another manner. Although a complete reranking would ideally have been required, time requirements of this procedure made it impracticable. Therefore, another method was adopted. A previous study by Calvin and Holtzman (1953) suggested that the reliability of S scores in pooled rankings is at least .90. As a check on this finding, a group was selected at random and requested to make a second set of rankings a month later. The correlation between initial and second ranking for the 15 members was .89. In view of the high reliability despite the passage of time, it seems conservative to assume a reliability of .90 for self-rankings for the total population of subjects. To reduce the greater sampling fluctuation of the S score as compared to the G score, the standard deviation of self-rankings also was computed for the total population of subjects rather than for each individual group.

The chance of selecting an extreme subject incorrectly was to be no greater than .01. Hence, the criteria for selection of the hostile-insightful group were $G \geq 2.58$ (standard error of measurement of G) + mean of G, and $S \geq 2.58$ (standard error of measurement of S) + mean of S. For the friendly-insightful group a similar formula was used except that now the cutting score was below the mean. The criteria were, therefore, $G \leq M_g - 2.58\,SE_g$, and $S \leq M_s - 2.58\,SE_s$, where SE and M refer to standard error and mean, respectively. Similar logic was followed in selecting the hostile-noninsightful and friendly-noninsightful groups. For the first of these groups, G was $\geq M_g + 2.58\,SE_g$. To insure noninsightfulness, the self-ranking had to be sufficiently below the G ranking so that the risk of error was to be .01 or less. Hence, the G−S discrepancy had to be $\geq 2.58\sqrt{SE_g^2 + SE_s^2 - 2r_{gs}\,SE_g\,SE_s} + M_{(g-s)}$. For the friendly-noninsightful group, G was $\leq M_g - 2.58\,SE_g$. The amount by which an individual's S score had to exceed his G score for the individual to be noninsightful was $\geq 2.58\sqrt{SE_g^2 + SE_s^2 - 2r_{gs}\,SE_g\,SE_s} + M_{(g-s)}$.

Table 1 summarizes the cutting scores for all four groups.

The average reliability of the G scores for the 25 student groups was .86, the range being from .50 to .94. The standard deviation (SD) range of G varied from .57 to 1.21, while the SEs ranged from .24 to .64. The reliability of the S scores was estimated to be .90, and the SD and SE values computed for the total population were 1.79 and .56, respectively. These S values were used as constant S criteria for the selection of subjects from each of the 25 groups.

Since the objectivity of perception of a particular group member was measured in terms of the M of the group judgments for him, and since each experimental subject had contributed to the determination of the objectivity score, the G score was recomputed, omitting in turn each subject's judgment score for each individual in his group. Then each subject's judgment for each member of his fraternity was compared (without regard

TABLE 1
CRITERIA FOR SELECTION OF FOUR PERSONALITY GROUPS

		Self-Rankings for Hostility	
		High	Low
		Hostile-insightful	Hostile-noninsightful
Group Rankings For Hostility	High	$G \geq M_g + 2.58\ SE_g$ $S \geq M_s + 2.58\ SE_s$	$G \geq M_g + 2.58\ SE_g$ $(G - S) \geq 2.58 \sqrt{SE_g^2 + SE_s^2 - 2r_{gs}SE_gSE_s} + M_{(g-s)}$
	Low	$G \leq M_g - 2.58\ SE_g$ $(S - G) \geq 2.58 \sqrt{SE_g^2 + SE_s^2 - 2r_{gs}SE_gSE_s} + M_{(g-s)}$	$G \leq M_g - 2.58\ SE_g$ $S \leq M_s - 2.58\ SE_s$

to sign) to the M judgment for that individual by the remaining members to yield a discrepancy score. The M of these discrepancies for all individuals in the group represented the individual judge's inaccuracy score. The higher the discrepancy the greater the inaccuracy score.

RESULTS

In correlating G and S scores for each of the 25 groups, the range was found to be from −.33 to +.71. The average r obtained via Fisher's z transformation was .18 ($p > .05$). For the group as a whole, therefore, the subjects self-perception scores bear little relationship to the peer judgments.

To test Hypothesis 1 that "hostile" subjects would perceive others more inaccurately than would "friendly" subjects, a 2 × 2 factorial design was set up for the dimensions "group judgments" (hostile or friendly) and "self-judgment" (hostile or friendly). Inspection of the plotted scores for each treatment combination revealed no gross departure from normality of distribution, and Bartlett's test for homogeneity of variance indicated that the four groups did not differ significantly with regard to size of variance. The results of the analysis of variance are shown in Table 2.

In accord with the hypothesis, there is a significant difference between the "hostile" and "friendly" groups in accuracy of perception of hostility in others, with the "hostile" group being more inaccurate. Differences in accuracy were not significant with respect to self-judgments or the interaction of group and self-judgments.

TABLE 2
ANALYSIS OF VARIANCE OF PERCEPTUAL INACCURACY SCORES FOR THE TRAIT OF HOSTILITY

Source of Variance	*df*	*MS*	*F*
Between Group Judgments	1	.327	7.27
Between Self-Judgments	1	.015	.33
Interaction Group × Self	1	.101	2.24
Within Groups	76	.045	

Note.—For 1 and 70 *df*: 7.01 = significant at .01 point, 3.98 = significant at .05 point.

The mean score for each group is shown graphically in Figure 1. The differences between the groups do not appear very large. It must be remembered, however, that in any comparison between the groups and single judgments, there is a considerable regression on G towards the mean as a result of the number of judgments comprising the G score. This is not true of the single judgment. It follows, therefore, that even for the most accurate perceiver, a considerable disparity will exist between the G score and his own judgment. The question is then whether the perceptions of any of the groups are relatively more objective.

Each group, therefore, was compared with every other group for differences between means (Table 3). Both the hostile-insightful and hostile-noninsightful groups were significantly more inaccurate than the friendly-insightful group ($p < .05$ and $p < .01$, respectively). A comparison of the insightful group (hostile-insightful, friendly-insightful) versus the noninsightful group (hostile-noninsightful, friendly-noninsightful) resulted in a significantly greater mean for inaccuracy on the part of the noninsightful group ($p < .01$). Hypothesis 2 is therefore supported.

The generality of these results might be questioned if they could be shown to depend on a few fraternities that contributed large numbers of subjects to the four groups. The

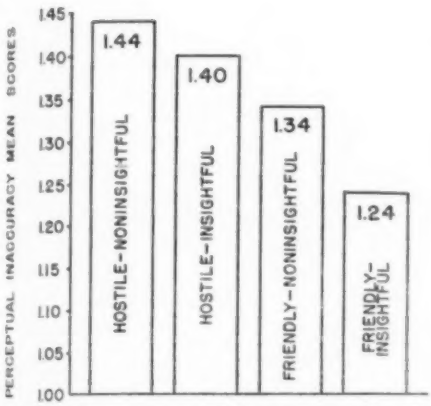

FIG. 1. Mean perceptual inaccuracy scores for the four extreme personality groups.

fraternity origin and score of each subject was therefore examined. Twenty-three of the 25 fraternity and dormitory groups supplied at least one subject. Four groups supplied six subjects; 2 groups, five subjects; 1 group, four subjects; 10 groups, three subjects; 5 groups, two subjects; and 1 group, one subject. Each fraternity group was then investigated to determine whether the general rank order of the means with regard to perceptual inaccuracy, as listed in Figure 1, would be found. The mean perceptual inaccuracy score for each personality grouping was compared with that for each other group that was present within the fraternity. A total of 37 comparisons were thus obtained from all of the fraternities. According to chance, 18.5 of these should be in the right order according to the ranking of the means and 18.5 in the wrong order. Actually, 27 were in the correct order and only 10 were incorrect. The likelihood of this result occurring by chance, as tested by χ^2, was less than .01, lending considerable support to the generality of the findings for the fraternity groups investigated.

TABLE 3

t VALUES BETWEEN ALL PERSONALITY GROUPS AND BETWEEN INSIGHTFUL AND NONINSIGHTFUL GROUPS FOR ACCURACY OF PERCEPTION OF HOSTILITY

Groups	t
Hostile-insightful versus Hostile-noninsightful	.66
Hostile-insightful versus Friendly-insightful	2.31*
Hostile-insightful versus Friendly-noninsightful	.85
Hostile-noninsightful versus Friendly-insightful	2.97**
Hostile-noninsightful versus Friendly-noninsightful	1.51
Friendly-insightful versus Friendly-noninsightful	1.46
Insightful Group (Hostile-insightful, Friendly-insightful) versus Noninsightful Group (Hostile-noninsightful, Friendly-noninsightful)	3.02**

* Significant beyond .05 point.
** Significant beyond .01 point.

DISCUSSION

The data seem to warrant the conclusion that, at least for the fraternity men studied in the sample, extreme possession of the trait of "hostility" significantly distorts the accuracy of perception of this trait in others. The evidence also indicates that insightful persons of extreme hostility or friendliness are significantly more objective in their perceptions than noninsightful persons. The effect of insight itself, however, probably is secondary to the absence of the trait of hostility in insuring objective perception. This statement is supported by the results of Table 3 which show that the t values for the hostile-noninsightful versus hostile-insightful groups and for the friendly-noninsightful versus friendly-insightful groups did not reach significance at the .05 level though they suggest a trend indicating that insight does facilitate accuracy of perception. It is only when both insightful groups are combined and compared with the combined noninsightful groups that the trend reaches significance. On the other hand, the subgroups compared for absence or presence of hostility with insight held constant yielded a significant t value ($p < .05$) for the hostile-insightful versus friendly-insightful group, and a near significant value for the hostile-noninsightful versus friendly-noninsightful groups.

The self-ranking apparently had no effect on the perceptual judgments, although the method of measuring the self-concept does not seem as valid (though it may be as reliable) as the measurement of the objective possession of the trait. When a large group of individuals reliably judge "A" to be 2.5 σ above the M for "hostility," and by ranking we eliminate "halo" and "leniency" effects, we may have some confidence that "A" is hostile at least in that behavior observed by his fraternity brothers. Such is not the case with self-judgments.

One is less likely to be honest with others

and/or with oneself in estimating self-possession of "hostility." Consequently, the self-rankings should be heavily influenced by personal needs. Some subjects may rank themselves as very hostile according to the code of "see no evil in thy fraternity brothers." Others may be very defensive about their possession of hostility and rank themselves as being low in hostility. The effect of varying "sets" should result in a comparatively large variance of self-judgments. The evidence bears this out in that the *SD* of S pooled for greater stability for the entire population was 1.79, as compared to the range of .57 to 1.21 for the *SD* of G for the different fraternities. It is apparent then that less confidence exists as to the validity of self-judgments as compared to group judgments.

SUMMARY

Four groups of 20 college men (hostile-insightful, hostile-noninsightful, friendly-insightful, friendly-noninsightful) were evaluated for accuracy of judgments of their fraternity brothers, using the pooled-rank method. The normalized mean ranking was used as the criterion and the mean discrepancy between a given individual's judgments of his fraternity brothers and the value assigned them by the rest of the group constituted the perceptual inaccuracy score. Results indicated that the groups differed significantly in accuracy. The rank order of the groups for accuracy of perception from most to least accurate was friendly-insightful, friendly-noninsightful, hostile-insightful, and hostile-noninsightful. Significant differences in accuracy of perception were found between hostile and friendly persons and between insightful and noninsightful ones. No significant differences were found between individuals who judged themselves to be friendly and those judging themselves as hostile, nor for the interaction between group and self-judgments.

REFERENCES

CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and the discrepancy between self concept and the inferred self. *J. consult. Psychol.*, 1953, **17**, 39–44.

GUILFORD, J. P. *Psychometric methods.* (2nd ed.) New York: McGraw-Hill, 1954.

MURSTEIN, B. I. The projection of hostility on the Rorschach and as a result of ego-threat. *J. proj. Tech.*, 1956, **20**, 418–428.

(Received January 11, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 221-230

AGE DIFFERENCES:

A DEVELOPMENTAL STUDY OF DISCRIMINABILITY AND AFFECTIVE RESPONSE[1]

NATHAN KOGAN[2]
Educational Testing Service

JUDITH W. STEPHENS
Cambridge, Massachusetts

AND FLORENCE C. SHELTON
Simmons College

The present paper addresses itself to the following two questions: First, to what extent do young children accurately discriminate age differences? Second, what is the relationship between the child's capacity to discriminate age and his preference and esthetic judgments of people varying in age? The first question is obviously concerned with perception and concept formation, the second with affective and attitudinal states in the child.

Consider first the perceptual-cognitive aspects of the problem. There is evidence that children at a very early age are capable of physiognomic discrimination. The well-known crying response to the stranger generally occurs well before the child is one year old. By contrast, the development of various categories or concepts for identifying or classifying people occurs considerably later. It is one of the purposes of the present study to specify when in the life of the child the age concept is acquired, as defined operationally by the child's capacity to make accurate age discriminations.

The second aspect of the study concerns the esthetic and social consequences of age discriminability in young children. Consider first the case in which the child has little awareness of age differences. Since such children are nevertheless able to distinguish between people, they can be expected to find certain people more attractive than others, and perhaps, as a consequence, like these people more than the others. It would be most unexpected, however, to find any basis in age for such differential preference and esthetic judgments. A systematic bias (positive or negative) with regard to a particular class of people seems unlikely before the child has an adequate concept of the class in question.

Consider next those children who can discriminate age differences. When this stage of cognitive development is reached, children might well begin to exhibit certain preferential tendencies based on the age of people encountered. If a bias against old people is manifested, the present study might have something to say about the origins of stereotypes often attributed to old people in American society (Tuckman & Lorge, 1953).

The present investigation is evidently analogous in many respects to the recent work on the development of racial attitudes in children. Studies by Goodman (1952) and Landreth and Johnson (1953) have shown that accurate perceptual discrimination of Negroes and whites by young children is often accompanied by a decided tendency to prefer whites to Negroes. It is our impression that the emphasis on the racial problem has been accompanied by a corresponding neglect of other important attributes of people in studies of children's social perceptions. We consider age to be one such important attribute and, hence, have undertaken to explore its impact on children in a photographic judgment context.

[1] This study was supported by the National Institute of Mental Health, United States Public Health Service, under Research Grant M-1867, and was conducted under the auspices of the Age Center of New England, Inc. A preliminary version of this paper was presented at the April 1959 meetings of the Eastern Psychological Association.

[2] Grateful thanks are due Peggy Golde for aid in the design of the study and Robert Elashoff for statistical advice.

METHOD

Subjects

A total of 102 children, 53 boys and 49 girls ranging in age from 4 to 6, served as the subjects of the study. For the purposes of the present study, subjects aged 4-0 through 4-12 were considered 4-year-olds. Corresponding age categorizations were employed for 5- and 6-year-olds. Subjects were obtained from summer day camps and private schools in the Boston area.[3] The fathers of the large majority of the subjects were in professional and managerial occupations. Hence, the present sample can be designated as predominantly middle to upper middle class.

Procedure

The stimulus materials consisted of two sets of passport photographs chosen from a pool of more than 100 photographs on the basis of similarity in emotional expressiveness and general "pleasantness" of appearance. One set contained photographs of nine females for whom mean age estimates made by a group of 20 adult judges ranged from 9.4 to 75.7 years. The other set consisted of photographs of nine males whose mean ages, as estimated by the same adult judges, ranged from 11.2 to 76.1 years. Age distances between photographs were of approximately equal magnitude in both sets. Agreement between the adult judges' age estimations was very high. Kendall's coefficient of concordance applied to the judges' age rankings of the two photograph sets yielded *W*s of .99 and .98 for male and female photographs, respectively.

Subjects were seen individually by a female experimenter. After the experimenter established some rapport with the subject, she would spread one of the picture sets before him (her). The child was then told to look the pictures over carefully and to separate them into two piles—those liked and those not liked. Within each pile the subject chose the picture liked best, next best, and so on, until each of the nine pictures had been assigned a rank. After this preference ranking was completed, the pictures were reshuffled, and the same procedure was repeated for the criterion of attractiveness—which is the best looking or the prettiest, the next best, and so on? Finally the subjects ranked the pictures on the basis of age—which do you think is oldest, next oldest, and so on?

The second set of photographs was then spread before the child, and preference, attractiveness, and age rankings again obtained. The particular order of the three questions—affective judgments preceding age judgments—was intended to insure that most subjects could cope with the task at its inception. If age judgments preceded those of preference and attractiveness, subjects unfamiliar with the age concept might well lose interest.

Half of the subjects received the set of male photographs first; the other half first responded to the female set. Upon the subject's completion of the task, a brief interview was conducted to provide additional information about the subject's awareness of signs of age and feelings about age differences. The following questions were asked: How do you know when someone is old? How do you know when someone is young? Whom do you know that you think is old? Are your parents young or old? Are your grandparents young or old?

Scoring categories were derived for each question. Subjects' responses were coded independently by two judges with agreements ranging from 91 to 96% for the five questions. Discrepancies were resolved by a third judge.

RESULTS

Since the basic data of the study are in the form of ranks, Spearman's rho was used to determine the magnitude of the relations between the variables. The mean age ranks of the 20 adult judges constituted the standard against which the children's age, preference, and attractiveness judgments were matched. The median correlations are shown separately for male and female subjects in Figures 1 and 2. For each figure, the results have been plotted separately for the male and female photograph sets. Table 1 presents the quartile deviations associated with the median values.

Age Discriminability

The four age discriminability curves (the solid lines of Figures 1 and 2) are distinguished, in general, by a substantial increase in discriminability from ages 4 to 5, with little further increase noted from ages 5 to 6. All of the median values are of moderate to high magnitude. Note, further, that the quartile deviations shown in Table 1 for subjects' age rankings are inversely related to subjects' age. The relatively large quartile deviations for 4-year-olds in relation to the moderately sized medians for that group indicate that some of the 4-year-olds are capable of performing the task as well as their age seniors while others in this age group are incapable of accurate age discriminations.

By the fifth year, the large majority of the present sample of children closely approximate adult performance in making relative age discriminations. Median values are quite high, and the quartile deviations (Table 1) are generally of low magnitude. With the 5-year-olds

[3] We wish to thank the following persons for making subjects available: Herbert J. Callahan of Mt. Ida Day Camp; Louise Keller of Page School, Wellesley College; Judith A. Schoellkopf of the Harvard Preschool; Decker Schwannecke of Dorchester House; Mary V. Wright of the Lesley-Ellis School.

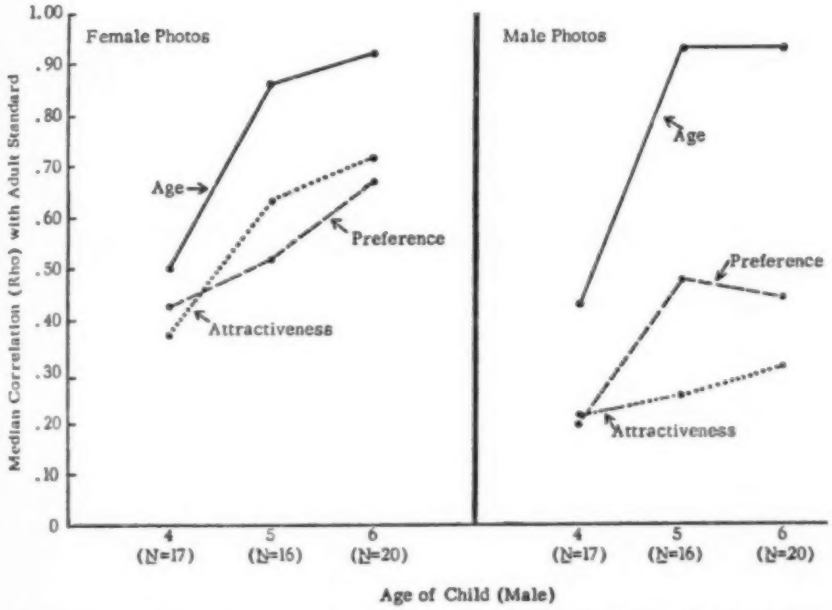

FIG. 1. Median correlations (Spearman rho's) between adult age rankings and male children's age, preference and attractiveness rankings.

proving so very capable at the task, any substantial further increase in age discriminability at the 6-year-old level is precluded. For the present sample of children, then, 5 is the age at which ordinal age judgments begin to conform to the adult criterion. This does not imply, of course, that 5- and 6-year-old children approximate adult performance in making absolute chronological age estimates.

Regarding sex differences in the capacity to discriminate age, the results indicate a slight superiority for the girls at the 4-year-old level for both sets of photos. These sex differences, however, do not achieve statistical significance (Mann-Whitney U test).

Differences in age judgments between photo sets are shown in Table 2. These were obtained by subtracting each subject's correlation for male photos from his (her) correlation for female photos. Since the distribution of these differences did not manifest any systematic positive or negative directional bias, the sign of the difference was ignored in obtaining median and Q values. Table 2 shows that the 4-year-old female subjects yield the largest median and Q values. On the whole, however, these values are quite small given the theoretically possible magnitude of intrasubject discrepancies. Apparently, age cues are of approximately equal salience for male and female stimulus faces. From another perspective, these results may be considered indirect evidence of the high reliability of age judgments.

Relation of Age Judgments to Preference and Attractiveness Judgments

To determine whether a child's preference and attractiveness judgments were biased toward the younger or older stimulus figures, the child's rankings were correlated with the mean adult age rankings. This procedure was selected over the alternative of correlating the preference and attractiveness judgments with the child's own age rankings, for the meaning of the latter correlations would depend upon the degree to which the child accurately discriminated age. The adult standard, on the other hand, serves as an appropriate common baseline for all subjects.

The median rho's for preference and attractiveness rankings are shown by age of child in the dashed and dotted lines, respectively, of

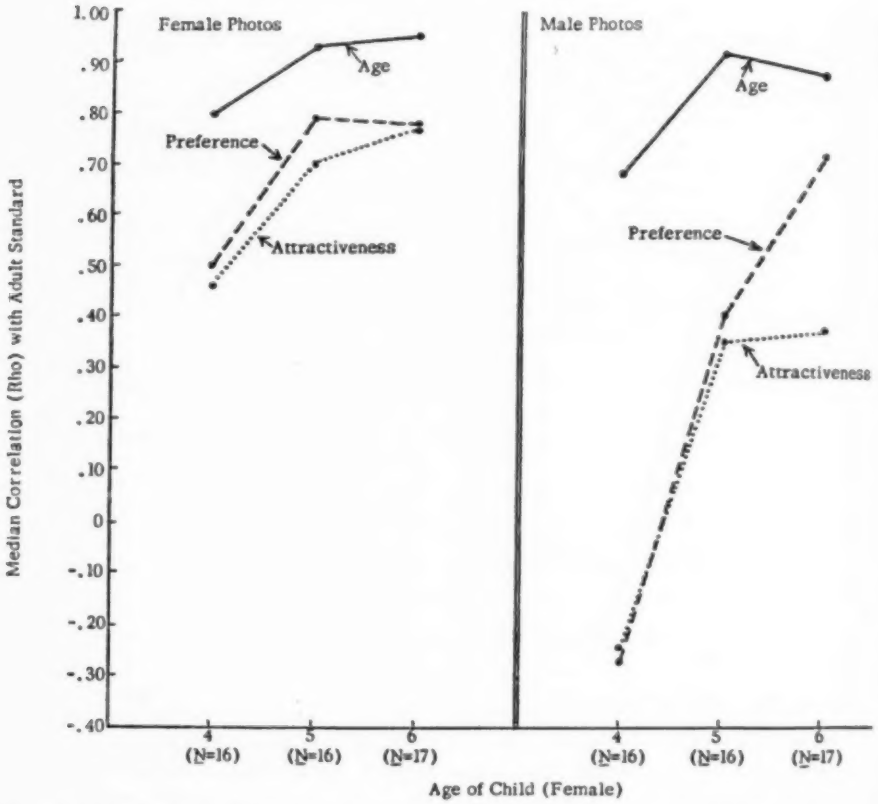

FIG. 2. Median correlations (Spearman rho's) between adult age rankings and female children's age, preference, and attractiveness rankings.

Figures 1 and 2. A positive correlation indicates a stronger preference for and more attractiveness attributed to the younger stimulus person. It will be noted that the preference and attractiveness curves are reasonably close to one another for both sex of the subject and photograph, and, to varying degrees, below the age discriminability curves. These data strongly suggest, then, that an age biased affective response to the photographs does not precede the stage of cognitive development required for accurate age discriminations.

Variability (quartile deviations) of the individual preference and attractiveness correlations are reported in Table 1. In contrast to the values for age reported in the same table, the variability of the preference and attractiveness rho's, in general, shows a less steep decline with increasing age of the subject. There is considerably more agreement, then, regarding who is younger and older than there is with respect to judgments of liking and attractiveness of the same photographs. This finding is hardly surprising, of course, for while judgments of age are quite stimulus-bound, preference and attractiveness judgments involve attitudinal and personality factors to a considerably greater extent.

Returning to Figures 1 and 2, it will be noted that affective responses to the female photograph set reveal, for both male and female subjects, an increasing bias in favor of the younger stimulus figures over the 4- to 6-year age span. (The preference curve for female subjects reaches its peak at age 5.) Subject sex differences are more pronounced for the male

TABLE 1

QUARTILE DEVIATIONS OF DISTRIBUTIONS OF CORRELATIONS BETWEEN THE SUBJECTS' RANKINGS AND ADULT STANDARD

Age and Sex of Subject		N	Female Photos			Male Photos		
			Age	Preference	Attractiveness	Age	Preference	Attractiveness
Male	4	17	.22	.32	.29	.36	.32	.39
	5	16	.16	.31	.22	.07	.27	.37
	6	20	.06	.18	.19	.09	.21	.33
Female	4	16	.33	.34	.30	.24	.31	.35
	5	16	.07	.18	.19	.06	.34	.25
	6	17	.05	.17	.17	.07	.30	.30

TABLE 2

MEDIANS AND QUARTILE DEVIATIONS FOR ABSOLUTE DISCREPANCIES BETWEEN MALE AND FEMALE PHOTOS IN CORRELATIONS OF THE SUBJECTS' AGE RANKINGS WITH ADULT AGE RANKINGS

Age and Sex of Subject		Median	Q
Male	4	.08	.11
	5	.10	.08
	6	.04	.08
Female	4	.19	.21
	5	.04	.07
	6	.10	.06

photograph set. The difference is largely one of curve slope, female subjects showing a substantial increase from age 4 to 6 in their preference for the younger stimulus persons, while males manifest a less marked increase in preference, and, indeed, show a slight decline at the 6-year-old level. Note further that the attractiveness curves deviate more from the preference curves for the male as opposed to the female photos. This effect very probably reflects the child's increasing awareness of cultural values which link appearance and desirability more strongly for females than for males.

Consider next whether the subjects manifest consistent preference or attractiveness biases for male or female photos. To provide an answer to this question, each subject's preference and attractiveness correlations for male photos were subtracted from his (her) corresponding correlations for female photos. Thus, a positive discrepancy indicated the subject's preference and attractiveness judgments were more strongly related to the adult age baseline for the female photos, while a negative discrepancy indicated that these relationships were stronger for the male photos. In every case (4-, 5-, and 6-year-old male and female subjects on preference and attractiveness—12 comparisons in all), the sum of the ranked positive discrepancies exceeded the sum of the ranked negative discrepancies. In other words, preference and attractiveness judgments were uniformly more likely to favor the younger female stimulus faces than the corresponding male faces.

A final question asked of the photograph judgment data was the extent to which the capacity to discriminate age influenced preference and attractiveness judgments of stimulus persons varying in age. To answer this question, the sample was arbitrarily divided into discriminators and nondiscriminators using as the cutting point a rho between the subject's age and adult age rankings significant at the .05 level or beyond (a rho of .59, one-tailed, for $n = 9$ stimuli). Given this criterion for assignment, 81 subjects (79.5% of the sample) were classified as discriminators of the male photos, and 78 subjects (76.5% of the sample) were so classified for the set of female photographs.

Statistical significance at the .05 level was also chosen as the cutting point for an age bias in preference and attractiveness judgments. In the present case, however, a two-tailed criterion was employed, since liking may be associated with either end of the age continuum. Accordingly, those subjects yielding Spearman rho's exceeding $\pm.68$ were considered to be manifesting age-biased preference and attractiveness judgments.

The relationships between age discrimina-

TABLE 3

PERCENTAGE OF THE SUBJECTS FAVORING YOUNGER AND OLDER STIMULUS PERSONS (Significant, Positive, and Negative rho's for Preference and Attractiveness with Adult Age) BY CAPACITY TO DISCRIMINATE AGE

Age Discrimination	Photo Set	Preference		Attractiveness	
		Favor Younger	Favor Older	Favor Younger	Favor Older
Yes ($N = 78$)	Female	56.4	0.0	60.3	0.0
No ($N = 24$)	Female	4.2	4.2	8.3	0.0
Yes ($N = 81$)	Male	35.8	3.7	32.1	4.9
No ($N = 21$)	Male	9.5	9.5	4.8	9.5

bility, on the one hand, and preference and attractiveness, on the other, are shown in Table 3. For both sets of photographs, a very small percentage of the nondiscriminating subjects shows an affective bias in favor of either younger or older faces. On the other hand, the subjects who accurately discriminate age are much more favorably disposed toward the younger stimulus persons. Of those subjects who do not favor the youngest faces in their preference and attractiveness judgments, only a small percentage are inclined in favor of the oldest faces in the photograph sets. This finding suggests that the remaining subjects may show a systematic bias toward faces in the middle-age range, or what is more likely, the judgments of these subjects may be based on attributes of the stimulus pictures other than age.

Responses to Interview Questions

Are there any relationships between the subject's responses to the interview questions and his (her) capacity to discriminate age? Before examining this problem, further consideration must be given to the division of the sample in terms of age discriminability. The distinction between discriminators and nondiscriminators shown in Table 3 was reported separately for the male and female photograph sets. If we wish to characterize a subject in terms of the accuracy of his age judgments on *both* photo sets, the sample must be divided in a different way. In brief, a total of 72 subjects satisfied the age discriminability criterion (rho = .59) on *both* photo sets. Of the remaining 30 subjects, 15 subjects met this criterion on *either* the male or female photo set, and 15 subjects on *neither* set. In the presentation of the interview results, the distinction between the latter two groups of subjects will be maintained. The subjects who accurately discriminated age on only one set of pictures will be designated as "partial discriminators." Those subjects who fell short of the criterion on both photo sets will be called "nondiscriminators." At first blush, it might appear that the necessity for the above distinction is inconsistent with the data reported in Table 2. The use of any arbitrary criterion to divide a distribution, however, often results in some scores relatively close to one another being relegated to different classes.

We turn now to the interview data. The subjects' answers to the question—"How do you know when someone is old?"—have been categorized in Table 4 for the subjects who meet, partially meet, and do not meet our criterion of age discrimination. Not surprisingly, the discriminators exceed the partial and nondiscriminators in reference to physiognomic and other old-age-related characteristics. The partial and nondiscriminators, on the other hand, are more likely to mention body size or, in the case of the partial discriminators, other physical characteristics unrelated to old age. Note, finally, that approximately one-half of the nondiscriminators are unable to specify any basis for judging age.

TABLE 4

RESPONSES (in Percentages[a]) OF DISCRIMINATORS, PARTIAL DISCRIMINATORS, AND NONDISCRIMINATORS TO THE QUESTION: "HOW DO YOU KNOW WHEN SOMEONE IS OLD?"

Response Category	Discriminators (N = 72)	Partial Discriminators (N = 15)	Nondiscriminators (N = 15)
A. Old-age-related physiognomic characteristics (e.g., wrinkles)	62.5	20.0	0.0
B. Other old-age-related characteristics (e.g., walk slow)	33.3	6.7	13.3
C. Bodily size (e.g., big, large, tall)	18.1	40.0	40.0
D. Other physical characteristics unrelated to old age (e.g., wear glasses, use make-up)	16.7	33.3	6.7
E. Response associated with death	2.8	13.3	0.0
F. Undesirable personality qualities and/or behavior	4.2	0.0	6.7
G. Unable to specify	9.7	20.0	46.6

[a] The total percentage in each column exceeds 100.0 since the total number of responses exceeded the number of subjects. There were also slight variations in the over-all number of characteristics cited in each group. Statistical tests were not attempted for these reasons.

Responses[4] to the question "How do you know when someone is young?" were consistent

[4] Four tables giving the response categories and response percentages for discriminators, partial discriminators, and nondiscriminators on Interview Questions 2, 3, 4, and 5 have been deposited with the American Documentation Institute. Order Document No. 6746 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

with these results, discriminators most frequently mentioning the absence of the physiognomic qualities attributed to the "old" in the previous question, partial and nondiscriminators referring primarily to bodily size or indicating that they do not know. Differences, however, were not as marked in the present case, for 28% of the discriminators, as opposed to 40% of partial and nondiscriminators, cite bodily size with reference to someone being "young." The concept "young," for some subjects, seems to connote a distinction between children and adults. In such cases, reliance upon physiognomic cues alone would obviously not be required for accurate discrimination between those "young" and "not young."

Consider next the question: "Whom do you know that you think is old?" Grandparents or persons of that generation were mentioned by 62.5, 33.3, and 13.3% of the subjects in the discrimination, partial discrimination, and nondiscrimination groups, respectively. Further, 46.7% of the partial and nondiscriminators were unable to specify anyone thought to be "old," as opposed to 18.1% for the discriminators. Of the various people who occupy the world of the young child, it is the grandparents who are likely to be the oldest. Hence, the between-group differences observed in the use of the "grandparent" category provide further validating evidence for the photographic judgment findings.

Finally, when asked the direct questions—"Are your parents young or old?" and "Are your grandparents young or old?"—subjects generally stated that parents are "young" and grandparents are "old." There were no substantial differences between groups in this regard, though such responses were most frequent for the age discriminators. Forty percent of the nondiscriminating subjects considered one parent "young," the other "old," but few subjects in the other two groups made this distinction. It appears most unlikely that the chronological age discrepancies between the parents of nondiscriminators substantially exceeded those of parents of the subjects in the other two groups. Accordingly, these results point to the confusion of the nondiscriminating subjects regarding the age dimension and suggest that many of these subjects make age distinctions on highly irrelevant grounds.

A few of the discriminators and partial discriminators (maximum of 15.3%) spontaneously categorized one or more of the parents or grandparents as "in between young and old." None of the nondiscriminators employed the above categories. It is entirely possible, of course, that the dichotomous quality of the two questions discouraged more frequent use of such a "middle-aged" response.

As for the partial discriminators, they resemble the discriminators on some of the categories, and the nondiscriminators on others. On still other categories, the partial discriminators manifest response frequencies about midway between those of the other two groups. There appears, then, to be a transitional phase through which some, if not all, children pass on their way to accurate age discriminability. At this stage, subjects seem to have an incipient awareness of age-relevant physiognomic cues, but have not as yet defined the age concept so as to disregard irrelevant stimulus information.

DISCUSSION AND CONCLUSIONS

The results of the present study indicate quite clearly that children as young as 4 years of age are capable of fairly accurate age discriminations solely on the basis of physiognomic information. Indeed, since a substantial proportion of the 4-year-old group was equal to the task, it is entirely possible that a small percentage of children in the late 3-year-old range would satisfy the discrimination criterion employed in the present experiment. It should be noted, however, that the present sample of children overrepresents the higher levels of socioeconomic status, and, very possibly, of intelligence, and the absolute level of age discriminability observed in the present sample may therefore well exceed the level that would be found in a truly representative sample of 4- to 6-year-old children.

Forty percent of the children who did not make accurate age discriminations on one or both photo sets, when asked in the interview "How do you know when someone is old?" responded in terms of the relative sizes of people. This finding can probably be traced to the fact that the child's first exposure to

the problem of aging and age differences centers around birthdays and the parents' emphasis on the child's being 2 or 3 years *old*, for example, accompanied by the assertion that the child is "getting to be a *big* boy or *big* girl now." Awareness of size differences occurs very early in the child's perceptual development, and, hence, age differences probably become linked to size in the eyes of many preschool children. There may, thus, be a stage of development in which children might be able to distinguish other children from adults without being able to discriminate adequately within the adult range. Children at this particular stage would be seriously handicapped, of course, in judging age from photographic stimuli. In time, the child can be expected to make finer differentiations within the adult category. When this stage is reached—by the fifth year for most of the children in the present sample—the child's perception is accurate in the sense that his categories of youth, middle, and old age match physiognomic and other cues associated with those particular age groupings.

It will be recalled that for the large majority of the nondiscriminating subjects, preference and attractiveness judgments were not systematically related to the ages of the stimulus figures. Of those subjects discriminating age accurately, on the other hand, approximately one-half manifest a clear preference for the youngest faces in the photograph sets. Correspondingly, the oldest stimulus persons are liked least. Attractiveness judgments tend to parallel preference judgments, suggesting that these are aspects of a global affective evaluation of the photographs.

While sex of the subject and of the stimulus person were of minor consequence in judgments of age, preference and attractiveness judgments, by contrast, were influenced to some extent by both sex of the subject and of the photograph. When comparing each subject's affective judgments of the male and female photos, it was found that judgments of the female photos were more directly age-linked, with the younger female faces viewed more positively. It is possible that a major determinant is the differential experience with adult males and females on the part of preschool children. Nursery schools and kindergartens almost exclusively employ female teachers, who quite naturally will vary in age. Whether older teachers are likely to be more "disciplinarian" in their educational practices[5] or simply are more remote in age from the subjects' mothers, the observed effect seems to be one of increasing negative affect with increasing age of the female stimulus person.

This explanation cannot, of course, account for the differential steepness of the male and female preference and attractiveness curves for the male photos. It will be recalled that the female subjects exhibited a sharp consistent increase with age in their positive affective response to the younger male photos. The curves for male subjects, on the other hand, showed a much more gradual increase in favor of the younger male stimulus faces, and, indeed, a slight decline between ages 5 and 6. We are uncertain as to how to interpret these data. One possible explanation would stress the growing awareness in the young female of emotional attachments to the opposite sex, an awareness constantly reinforced by the mass media. From this standpoint the youngest male photos would show an increase in the extent to which they are endowed with positive affect by female subjects. For male children, on the other hand, identification processes would naturally assume priority over object choice when male photos are the stimuli. Such a conceptualization would lead to the expectation that male subjects would prefer pictures of young and middle-aged adults to those of children and adolescents. Such a preference pattern would tend to reduce correlations between age and affective judgments.

This interpretation does not fully account, however, for the unusually low median correlations for affective judgments of the younger female subjects on the male photos. It is our opinion, on the basis of the subjects' spontaneous comments, that the response to the youngest member of the stimulus series is the critical factor in the present case. In the female photo set, the youngest face was considered a potential playmate by most of the youngest male and female subjects. The youngest face in the male stimulus series, on the other hand, met

[5] In a subsequent questionnaire study of 34 children aged 4 to 6, 25 (73.5%) expressed a preference for a "young" rather than an "old" teacher.

with a more equivocal reaction from the 4-year-old males, and with outright rejection from many of the youngest females. Possibly, the differential response is based on the relative ages of the two stimulus persons. It will be recalled that the adult judges, on the average, estimated the youngest male photos about 2 years older than the corresponding youngest female photos. The youngest male may, thus, have fallen outside the category of playmate. For the youngest female subjects, in addition, the male photo in question probably called forth an undesirable image of "aggressive masculine play." As we have seen, such an image, if present, is not maintained, for the 5- and 6-year-old female subjects.

In conclusion, we should like to consider some of the more general social forces that might contribute to the emergence in young children of negative sentiments toward old people. First, however, it should be recalled that the study employed ranking as its basic procedure. While ranked judgments have the decided merit of comprehensibility for young children, they do have the disadvantage of failing to provide any absolute yardstick in regard to affective responses to the photographs. Low rankings may simply be reflections of indifference or, in the rare case, of mild positive feelings.

Considerations such as these are especially relevant if one resorts to a "role" interpretation of the results. Thus, if the youngest stimulus persons are perceived relative to their role as playmates, a highly valued role for 4- to 6-year-old children, the youngest figures can be expected to receive the most positive endorsements. Faces in the young adult and middle age range may be perceived as fulfilling parental and teacher roles, both not nearly so pervasively positive as the playmate role. Faces in this age range, then, can be expected to receive a lower ranking. Finally, there are the oldest stimulus persons in the photograph series, for whom a clear role is entirely lacking. These faces perhaps tend to receive the least positive affective judgments by default.

But how about the grandparents of these children? The interview data indicated that grandparents are generally identified as "old." Insofar as grandparents constitute an indulgent rewarding figure for children, one might expect some generalization to the oldest faces, and, hence, more favorable judgments of old people. It is evident, however, that the relation between grandparents and the young child in our society is a peripheral one, relative to the frequent intensive contact with parents and peers. Nevertheless, a positive emotional relationship with one or more grandparents might serve to mitigate negative feelings toward old people, and, in the present study, might partially account for the fact that a substantial number of discriminating subjects did not assign their lowest preference ranks to the oldest stimulus figures.

Some consideration should be given to factors other than "role" that might contribute to a better understanding of the findings. Some children, for example, may develop negative attitudes toward old people as a consequence of the explicit or implicit transmission of stereotypes (Tuckman & Lorge, 1953) by parents or other adults. Such transmission may take the form of stories read to children. Elderly characters in classic children's tales are often cast in a stereotyped role, or are unsympathetically portrayed.

Anthony (1940) noted that 4- and 5-year-old children sometimes explicitly link death and old age. In the present study, only a few children *spontaneously* associated death and old age in the interview (Table 4). It is conceivable, however, that a large number of children were influenced by this connection, but failed to mention it in the absence of a direct question. While an association of this kind may apply to a limited number of children, it would serve to bias the affective response to the older faces in a negative direction.

Note, in conclusion, that only about half of the older children in the sample gave the oldest stimulus persons the lowest preference and attractiveness rankings. Why, then, did the other half not do so? We have suggested that a positive relationship with one or more grandparents might exert an influence. Are there any other special circumstances likely to mitigate the development of negative feelings in these subjects? Or, do children simply vary in the length of time intervening between age discrimination learning and the development of age-related affective predispositions? Clearly, these questions can be answered only

in terms of extending the present research to older children and inquiring further into factors responsible for individual differences in children's perception and evaluation of the age dimension.

SUMMARY

A sample of 102 children (53 males and 49 females) ranging in age from 4 to 6 was studied with regard to capacity to discriminate age differences and affective responses to stimulus persons varying in age. Each subject was presented with two sets of stimulus faces of heterogeneous ages with instructions to rank the photographs in terms of age, preference, and attractiveness. The stimuli, a male and female set each containing nine photos, had been ranked for age by a group of adult judges. This ranking served as a baseline against which the children's rankings were compared.

Progressive increases in age discriminability were noted from ages 4 to 5, setting a ceiling that permitted little further increase from 5 to 6. Affective judgments with increasing age were in the direction of a more positive response to the younger faces. Preference and attractiveness judgments were influenced to some extent by the sex of the subject and of the photograph set.

When the sample was divided into age discriminators and nondiscriminators, it was found that consistent preference and attractiveness judgments in favor of the younger faces were found only in the former subgroup. Evidence is reported indicating that partial and nondiscriminators are more likely to associate age with body size, while discriminators tend to use physiognomic cues.

Some possible interpretations of the findings are presented in the body of the paper.

REFERENCES

ANTHONY, SUSAN. *The child's discovery of death.* New York: Harcourt, Brace, 1940.

GOODMAN, MARY E. *Race awareness in young children.* Cambridge, Mass.: Addison-Wesley, 1952.

LANDRETH, CATHERINE, & JOHNSON, BARBARA C. Young children's responses to a picture and inset test designed to reveal reactions to persons of different skin color. *Child Develpm.*, 1953, **24**, 63–80.

TUCKMAN, J., & LORGE, I. Attitudes toward old people. *J. soc. Psychol.*, 1953, **37**, 249–260.

(Received January 12, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 231-241

COGNITIVE RISK AND ENVIRONMENTAL CHANGE[1]

JEROME S. BRUNER AND HENRI TAJFEL[2]
Harvard University *University of Oxford*

It is the objective of the present paper to examine wherein breadth of categorizing reflects the manner in which people deal with the risk of errors of judgment—specifically, the risk of saying that things are similar when they might be different, or that they are different when they might be similar.

Breadth of category—or, as it has sometimes been called, equivalence range—refers to the range of stimuli that are placed in the same class or category and share a common label. The criteria for placing objects in a common category may be many or few, explicit or implicit. In psychophysical experiments using unidimensional continua, such as length, breadth of category defines the limits beyond which stimuli are no longer classified as X, or "long," but either as non-X, "not long," or as Y, "short." In more complex arrays of stimuli, there is no such unidimensional simplicity, and the criteria for classification are often combined relationally or disjunctively in the process of deciding whether something does or does not belong in a class—as, for example, in classifying men as "competent" or "eligible to vote."

In recent years, individual differences in categorizing, particularly in breadth of categorizing, have been thought to reflect important general differences in cognitive functioning. Four trends of research on this problem can be distinguished: the investigation of changes in breadth of category as a function of changes in categorizing conditions, the study of individual consistencies in breadth of categorizing in different judging situations, the search for relationships between various emotional and personality factors and the individual's general preference for broad or narrow classifications, and the exploration of relationships between breadth of categorizing and various forms of abnormal mental functioning.

A word about each of these trends as they relate to the present study. It is apparent that when one alters the conditions of decision breadth of category changes. Thus, a sentry, deciding whether approaching figures in the twilight are friends or foes will have a broader inclusion category of "foes" than will an inspector judging machine parts as defective or not. For the consequences obviously vary in the two instances and one can control the severity of these consequences by altering breadth of category (see Bruner, Goodnow, & Austin, 1956). In general, it can be said that when the perceived consequences of overinclusion are more severe than the consequences of overexclusion, inclusion categories narrow. When overexclusion leads to more severe consequences, inclusion categories broaden.

We mention this line of research for two reasons. In the first place, it underlines the risk-regulating character of categorizing. But beyond that, it points to the possible meaning of consistencies in categorizing breadth that appear to be associated with different personality characteristics—the subject matter of the last three of the four lines of inquiry mentioned above. Thus, Gardner (1953), Hamilton (1957), Klein (e.g., 1958), and others remark that obsessive, anxious doubters tend to prefer narrow inclusion categories in their sensory judgments. Perhaps they have a tendency to minimize risk of error by the nay-saying route, preferring the consequences of error that come from avoiding contact with threatening objects. It has been shown, too, that the ambivalent, ethnocentric personalities studied by Frenkel-Brunswik (1949) and Rokeach (1951) show a tendency to narrow categorizing. Yet, on the other hand, Arnhoff (1956) has noted among ethnocentrics a tendency to overgeneralization—which is indeed what ethnocentrism implies. Interestingly enough, Arnhoff (1957) obtained completely negative results in an attempt to replicate his study. This negative replication, while one must be cautious in interpreting it, may be revealing in the sense that there may

[1] This research was done as part of a research project in cognition, supported in part by USPHS Grant M-1324.

[2] In 1958–59 in the Department of Social Relations, Harvard University, on leave of absence from the University of Oxford.

be two opposing tendencies operative in the ethnocentric personality: a tendency toward overgeneralization under certain conditions, and toward narrow categorizing under others. Zaslow (1950) suggests, and perhaps his suggestion is relevant in this connection, that normal subjects, in contrast to schizophrenics, tend to use middling wide categories, whereas schizophrenics veer either toward overnarrowness or overgeneralization.

In any case, it seems patent that experiments seeking to relate categorizing breadth either to conditions of judgment or to personality cannot get very far unless and until we understand much more about the intimate texture of the act of categorizing itself. This is not to say that we do not recognize the degree to which categorizing is affected by these two sets of determinants. Rather, we feel that an examination of the underlying judgmental processes involved in categorizing is necessary for a fuller understanding of the role played by motivational and personality variables.

We turn now to the series of experiments with which the present paper is concerned. Our central interest is in the specific behavior of broad and narrow categorizers in a highly simplified judging task. Two questions concern us. The first has to do with consistency of preference for broad and narrow categorizing in a stable stimulus situation. But consistency is not as simple as it might at first appear. There is an equally important question concerning reactions to a changing stimulus environment. Is there some consistent manner in which broad and narrow categorizers alter their judgments in the face of changes in the stimulus situations with which they must cope?

With respect to the second question, some predictions can be made. If we assume that a narrow categorizer is one who is compelled to attend to differences in stimuli, whereas a broad categorizer tends to overlook differences, then it should follow that narrow categorizers will be more likely to *alter the breadth of their categorizing* to conform to changes in the stimulus world. Thus, narrow categorizers should show more susceptibility to anchoring effects in psychophysical judgment than do broad categorizers by virtue of being more sensitive or reactive to change and contrast. Indeed, we would say, it is precisely this sensitivity to change that constitutes the hedge against risk that characterizes the narrow categorizer. For to the narrow categorizer, differences and change are highly relevant, and, as Tajfel shows (1959), these relevent differences tend to be accentuated. And if this is the case, not only should the narrow categorizer alter his breadth of category more readily in response to stimulus changes, but he should also adapt more quickly when the change persists and becomes a new steady state of stimulation.

There is another way of stating our prediction, one taken from the language of anchoring experiments. When a subject has been judging a range of stimuli in terms of categories of absolute magnitude, the introduction of new stimuli larger than the range has either a "contrast" or an "assimilation" effect. A contrast effect is responding strongly to the newly-introduced stimulus—say a stimulus larger than any previously encountered—by judging previous stimuli as contrastingly smaller. Assimilation, both in common sense terms and technically, is "not heeding" the new anchoring stimulus, i.e., judging it as no larger than the stimuli previously encountered. We might well predict that narrow subjects, sensitive to differences, will be subject to contrast effects. Broad subjects, on the other hand, should by virtue of their less acute reactiveness to change, show less of the contrast effect and may indeed assimilate the stimuli of the extended range to the psychophysical scale established with the smaller range of stimuli.

With these predictions in mind, we may turn now to the experiments proper.

EXPERIMENT I

Subjects and Procedure

Forty-eight fifth grade school children served as subjects in this experiment. They were tested individually.

Stimuli consisted of projected slides containing clusters of dots, cast on a bended screen at an exposure time of .5 second per cluster. The subject, facing the screen 12 feet away, was seated next to the table on which the 35 millimeter Bell and Howell projector was placed. At the beginning of the session, each subject was told that his task would be to decide whether each of the clusters to be flashed on the screen contained 20 dots or did not contain 20 dots. If he thought that there were 20 dots in the cluster he was to say "yes"; if he

thought that the cluster did not contain 20 dots, he was to say "no." The subject was then told that 20 would be the smallest number of dots in any of the clusters that he would see, and was shown two sample clusters of 20 dots each, as demonstrations.

The presentation of the experimental series followed immediately. In order to avoid the possibility that the decision about the number of dots might be determined by the recognition of the shape of a cluster once seen, four clusters with different dispositions of dots in them were used for each number of dots, and, in addition, each of these slides was shown in different positions at its successive presentations.

The experimental presentations were divided into two parts. First, a series of clusters containing from 20-26 dots was presented eight times successively—a total of 56 presentations. The presentations were randomized in such a way that each cluster of the series appeared once before the whole series was repeated in a different random order. Following this, in the second part of a session, four series of clusters containing from 20-30 dots were presented in varying random orders—a total of 44 presentations.

The session was interrupted by two periods of rest of approximately 2 minutes, the first after three presentations of the series; the next, three presentations later.

The scoring was based on the percentage of yes responses, representing for each subject the proportion of the stimuli which for him fell into the category "20." These percentages were taken for each subject for three separate phases of the presentations of stimuli: for the initial range of 20–26 dots in its eight consecutive presentations (Phase I), for the stimuli 20–26 of the extended range in its first two presentations (Phase II), for the stimuli 20–26 of the extended range in its last two presentations (Phase III). Breadth of category was defined as the proportion of stimuli assigned by a subject to the inclusion category 20 in Phases I, II, and III, and a comparison of these phases provides us with comparative data for testing our hypotheses.

Results

Before the results are reported and discussed, the hypothesis stated in general terms earlier should be stated specifically in the context of the procedure just described. It was assumed that during the initial eight presentations of the stimuli in the original range, each subject would settle to his characteristic pattern of assigning a certain proportion of stimuli to the inclusion category of 20. The subjects could, thus, be divided, by means of ranking, into three groups: a group of broad categorizers with the largest proportions of number of calls of 20; a middle group, and a group of narrow categorizers with the smallest number of calls of 20.

Extending the series of stimuli from clusters ranging from 20–26 dots to clusters ranging from 20–30 dots was conceived as a form of anchoring, introducing a larger difference between the two extremes of the range. According to the argument outlined previously, this change should have been more salient and effective for the narrow than for the broad categorizers. In other words, after the introduction of the additional larger stimuli, the narrow categorizers should, in their responses to this part of the new range which constituted the original range (clusters from 20–26 dots), increase the proportion of clusters called 20 more than the broad categorizers. The narrow categorizers should, moreover, adapt back to their old narrow pattern with the continued experience of the new range provided by Phase III.

1. *Initial breadth of category.* The range in breadth of category for the original series of stimuli was wide: from 89.5%–30.3%, with the median proportion of calls of 20 being about 55%, with six subjects calling 40% or fewer 20, and eight calling more than 70% of the stimuli 20.

The division of subjects into three groups was made at "cut-off" points nearest to 27% of the distribution (Kelly, 1939) for the upper and lower ends of the distribution. This resulted in 12 subjects with the largest percentages of yes responses being assigned to the broad group, (with breadths ranging from 64.3%–89.5%), and 13 subjects with the lowest percentages of yes responses being assigned to the narrow group (with breadths ranging from 30.3%–44.6%). The difference between the number of subjects in the groups is due to a run of ties at the upper end of the distribution. Assigning all the remaining 23 subjects to the middle group seemed the least arbitrary procedure.

2. *Change in breadth of category as a function of extending the range.* As a first test of the general hypothesis, a rank correlation was run between breadth of category in Phase I and the shift from Phase I to Phase II in the percentage of calls of 20. This correlation was of the order of rho $= -.433$; with 46 *df* this is significant at $p < .005$. Thus, the larger the number of stimuli initially assigned to the category 20 by a subject, the less did he tend to increase this number for stimuli of 20–26

dots in the first two presentations of the extended range.

As Hamilton (1957) rightly points out, however, "there are few forms of behavior, especially in the cognitive sphere, which cannot be shown to vary with intelligence. In fact, it might be thought that intelligence is maximally involved in any process requiring perceptual judgments" (p. 203). In pursuit of this point, a rank correlation between IQ and breadth of category in Phase I was obtained. It was of the order of $-.53$. The initial breadth of category appears then to be determined, in part at least by the subjects' intelligence—more intelligent subjects tended to use narrower categories. The mean IQ of the broad group was 107.4; of the middle group, 113.6; and of the narrow group, 122.7. The difference between the broad and narrow groups is significant at $p < .002$.

The testing of predictions outlined above depended, however, on the relationships found between the breadth of category shown by the subjects in Phase I and the amount of shift from Phase I to Phase II. A rank correlation between the amount of shift and IQ was obtained, and found to be .27, reflecting the fact that narrow subjects shifted more. However, a partial correlation between IQ and shift, holding breadth of category constant, reduced this figure to .05. A complementary finding is that the partial correlation between breadth of category and amount of shift, this time holding IQ constant, remained as high as $-.35$. Thus, it would seem that, though breadth of category is consistently related to intelligence, the amount of judgmental shift that occurs when the environment changes is related not to intelligence but to breadth of category.

The Wilcoxon matched-pairs signed-ranks test (Siegel, 1956) was applied to assess the significance of the increase in the percentage of yes responses to stimuli 20–26 after the extension of the range for each of the groups. The results for the broad group were not significant. Out of 12 subjects in this group, 6 actually shifted in the direction of *decreasing* the proportion of their yes responses after the extension of the range. In the middle group the shift was significant at $p < .005$ (19 out of 23 subjects increased the proportion of their yes responses). In the narrow group, all 13 subjects increased their proportion of yes responses. The mean increase in percentages of yes responses from Phase I to II for broad categorizers was $+.45$, the range being from -21.4–$+21.4$. The comparable increase for the narrow categorizers was $+20.1$, with a range from $+3.6$–$+37.5$. The middle categorizers were in between: mean increase in percentage of yes responses being 12.3 with a range from -12.5–$+30.4$. Mann-Whitney U tests were then applied to compare the amount of shift found in the different groups. All of them were found to be significantly different from each other (broad vs. narrow, $p < .01$; broad vs. middle, $p < .02$; narrow vs. middle, $p = .05$).

It might also be said parenthetically that the broad and narrow groups behaved consistently in their placement of dot clusters ranging in number from 27–30 in Phase II. Some 40.6% of these were called 20 by the broad categorizers in contrast to 20.2% called 20 by the narrow categorizers, a difference significant at $p < .025$ by the U test.

3. *Changes in breadth of category as a function of adaptation to the extended range of stimuli.* Let us now compare the change from Phase II to Phase III—the subjects now having had experience with two presentations of the extended range and confronting the two additional series. In the narrow group, out of 12 subjects (data for this shift for one of the subjects were not available), 8 subjects shifted back towards fewer yes responses for the 20–26 part of the range, 1 did not shift, and 3 shifted towards more yes responses. The comparable figures for the broad group are interesting. Only 2 subjects in this group showed a decrease in yes responses in Phase III in comparison with Phase II; 7 subjects increased their yes responses and the remaining 3 subjects showed no shift at all. A comparison of means tells the same story. On the average, subjects in the broad group *increased* their yes responses by 7.0% between Phase II and III. Narrow categorizers *decreased* theirs 9.7%. The difference between the two groups tested by the Mann-Whitney U test is significant at a $p < .025$.

Here again there is a correlation of .37 between the initial breadth of category and the amount of shift from Phase II to Phase III,

indicating that broad categorizers tend to broaden their range more as they get more exposure to the changed stimulus situation, the narrow categorizers having done most of their changing when the stimulus situation was first altered. A correlation of −.29 was found to hold between IQ and the amount of shift, an obvious consequence of the fact that the more intelligent subjects had already done their shifting in Phase II. A partial correlation was computed between IQ and the amount of shift, from II to III, holding breadth of category constant. It was found to be −.12. But a partial correlation between breadth of category and amount of shift, holding IQ constant this time, is .27. It can be argued then, that the amount of shift from Phase II to Phase III, reflecting adaptation to the new range, is related to the initial breadth of category but not to intelligence, though intelligence and the initial breadth of category are related to each other.

The data for all subjects in the three phases of experiment are summarized in Table 1 in the form of percentage of yes responses given.

Figure 1 summarizes the findings. The main feature of the results is the relative stability of performance in the broad group and the sharp changes from phase to phase in the narrow group. This is congruent with the general conception of broad and narrow categorizers as outlined in this paper and in others (e.g., Gardner, 1953). The broad categorizers do not react to change markedly or rapidly; the small and insignificant amount of shift that they show after the extension of the range maintains a slow and steady increase as the new range continues to be presented. Narrow categorizers react to change strongly, but then they seem to revert rapidly to their preferred mode of response as soon as there is habituation to the new situation.

Our main hypothesis—that the narrow categorizers are more sensitive to change than broad ones—is confirmed.

There is one other point to be made in connection with cognitive risk. One can conceive of risk regulation in terms of a cost-for-gain reckoning. That is to say, for a gain in accuracy or security or some other "good," one pays a cost. In the case of the broad categorizers in Experiment I, we see that they do in fact call more of the stimuli containing 20 dots correctly. The cost of this is calling many more of the other dot clusters incorrectly. The narrow categorizers show a considerably sharper slope in their discrimination function. The price they pay for *not* being wrong about the excluded items is missing some of those

TABLE 1

PERCENTAGE OF YES RESPONSES GIVEN BY SUBJECTS IN THE THREE PHASES OF EXPERIMENT I AND IN EXPERIMENT III

Group	Experiment I					Experiment III
	Phase I (20–26)	Phase II (20–26)	Phase III (20–26)	Phase II (27–30)	Phase III (27–30)	
Broad						
1	89.5	71.4	78.6	62.5	62.5	58.9
2	85.7	100.0	100.0	37.5	75.0	64.3
3	83.9	78.6	100.0	0	37.5	69.5
4	80.3	64.3	85.7	37.5	50.0	71.4
5	80.3	92.8	100.0	100.0	100.0	100.0
6	75.0	71.4	85.7	25.0	37.5	75.0
7	71.4	50.0	78.6	62.5	50.0	94.6
8	71.4	78.6	50.0	62.5	75.0	58.9
9[a]	69.6					64.3
10	67.8	85.7	70.1	37.5	37.5	83.9
11	67.8	50.0	78.6	25.0	75.0	58.9
12[a]	64.3					62.5
13[a]	64.3					62.5
14	64.3	78.6	78.6	25.0	25.0	58.9
15	64.3	85.7	85.7	12.5	50.0	58.9
M^b	75.1	75.6	82.6	40.6	56.2	
M^d	73.3					69.5
Middle						
1	62.5	64.3	71.4	25.0	0	55.4
2	62.5	64.3	85.7	62.5	25.0	57.1
3	62.5	50.0	64.3	37.5	0	58.9
4	60.7	78.6	85.7	25.0	25.0	53.6
5	58.9	78.6	57.1	12.5	25.0	51.8
6	57.1	85.7	78.6	25.0	25.0	41.1
7	57.1	71.4	57.1	12.5	50.0	50.0
8	57.1	64.3	57.1	37.5	25.0	92.9
9	57.1	85.7	71.4	12.5	12.5	44.6
10	57.1	71.4	78.6	62.5	50.0	66.1
11	57.1	85.7	78.6	50.0	25.0	60.7
12	55.3	42.8	35.7	37.5	50.0	55.4
13	53.6	71.4	50.0	37.5	12.5	73.2
14[a]	53.6	78.6	71.4	37.5	12.5	
15[a]	51.8					57.1
16	51.8	50.0	50.0	37.5	12.5	48.3
17	51.8	64.3	85.7	25.0	62.5	41.1
18	51.8	71.4	64.3	0	12.5	57.1
19	50.0	64.3	78.6	37.5	12.5	50.0
20[a]	50.0					50.0
21	48.2	50.0	57.1	12.5	25.0	39.3
22	48.2	78.6	64.3	0	0	55.4
23[c]	46.4	57.1	35.7	0	37.5	57.1
24[c]	46.4	42.9	50.0	25.0	12.5	
25[c]	46.4	64.3	57.1	62.5	50.0	44.6
M^b	54.5	66.8	64.6	29.4	24.5	
M^d	55.3					55.2

TABLE 1—*Concluded*

Group	Experiment I					Experiment III
	Phase I (20–26)	Phase II (20–26)	Phase III (20–26)	Phase II (27–30)	Phase III (27–30)	
Narrow						
1	44.6	78.6		12.5		51.8
2	42.8	57.1	71.4	0	12.5	48.3
3	41.1	71.4	50.0	12.5	25.0	51.8
4	41.1	50.0	42.9	25.0	12.5	64.3
5	41.1	64.3	50.0	0	0	60.7
6[e]	41.1	78.6	78.6	87.5	87.5	
7	41.1	57.1	21.4	0	0	62.5
8	39.3	64.3	78.6	25.0	25.0	39.3
9[e]	39.3	42.9	35.7	0	0	
10	35.7	50.0	14.3	87.5	87.5	64.3
11	33.9	57.1	64.3	0	25.0	46.4
12	32.1	57.1	35.7	12.5	0	35.7
13	30.3	35.7	28.5	0	0	46.4
M[b]	38.7	58.8	47.6	20.2	22.9	
M[d]	39.7					51.8

[a] Subjects not included in the analysis of Experiment I, and added subsequently for comparison of their performance in Experiments I and III.

[b] Means for Experiment I, not including subjects marked a.

[c] Subjects taking part in Experiment I, and not available for Experiment III.

[d] Means for comparison of performance in Experiments I and III, not including subjects marked c.

[e] Subjects falling in the lower part of the distribution (narrow group) for comparison between Experiments I and III, and included in means for narrow group (d).

that should be included among the 20s. Broad subjects, in short, pay their price for being right; narrow subjects seem to pay theirs for not being wrong.

But before these, or any other conclusions are finally drawn from the results, some very obvious artifactual possibilities must be dealt with. The results may have been due not to the narrow group's more abrupt reaction to change, but to the fact that by definition the narrow categorizers have more room to expand their proportion of yes responses. A broad categorizer who started out with 80% of yes responses cannot, of course, achieve an increase of 25%. This "ceiling effect" may account for the broad group showing less increase in yes responses than the narrow group. Yet, the reversal of trends in shift from Phase II to Phase III argues in some measure against this artifactual explanation; so does the fact that half of the broad subjects *decreased* their proportion of yes responses from Phase I to Phase II, some of them quite considerably. The results, at least the shift from Phase I to Phase II it might also be argued, could have been due to regression effects. Experiment II was conducted to test further the validity of interpretation given to the results of Experiment I, particularly with respect to ceiling effects and regression effects.

Another artifact, possibly lurking in the results, is that the narrow group is "narrow" not because of its preferred mode of taking cognitive risks, but simply because, for one reason or another (such as, for example, in-

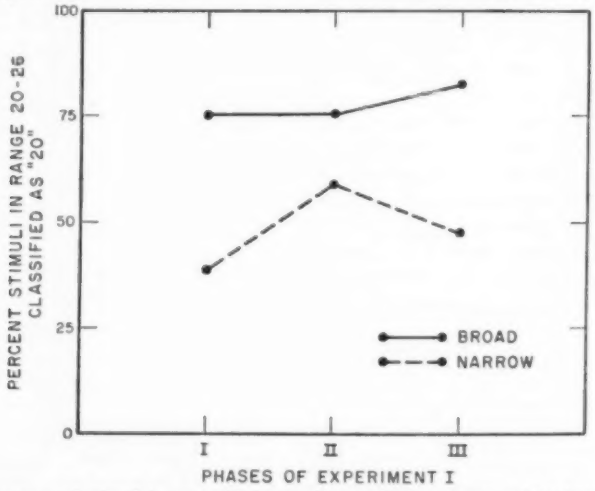

FIG. 1. Mean percentage of stimuli in range 20–26 classified as 20 (yes responses) by broad and narrow categorizers in the three phases of Experiment I.

terest in the task), the capacity of narrow subjects to discriminate was better than that of the broad group. The relationship between intelligence and breadth of category, previously reported, increases the likelihood of some such interpretation. Under the conditions of Experiment I and with the small number of yes responses that we obtained, the narrow group could by definition get closer to the actual leptokurtic distribution of the "correct" stimuli than the broad group. Thus, in Phase I, the proportion of clusters containing 20 dots was 14.3%, and decreased to 9.1% in Phases II and III. But note, however, that 7 out of 13 subjects in the narrow group *increased* their proportion of yes responses to the new range presented to them in Phase II, while only 3 out of 12 subjects in the broad group did so, a difference that is significant at the .02 level as estimated by a *U* test. There remains the possibility, however, that there is some intrinsic relationship between intelligence test scores and narrow categorizing based on the possible bias of intelligence tests in favor of the precise or narrow categorizers. Yet this is not borne out by the indifferent partial correlation between IQ and shift from Phase II to Phase III. But such indirect evidence is not sufficient. Consequently, Experiment III was conducted in order to explore further the question of accuracy.

EXPERIMENT II

In Experiment I, change was introduced by extending the series beyond its initial range. Using the same series of stimuli, it should be possible to shift judgments in the opposite direction by biasing the distribution of stimuli through *cutting off* a part of the range. The order of presentation used in Experiment I was reversed in Experiment II: The subjects were first confronted with the range of 20–30 dots, then the larger end of the series was eliminated and only the 20–26 range presented. If it is true that change has a greater effect on narrow than on broad categorizers, then in Experiment II the narrow categorizers should *decrease* their proportion of yes responses in the 20–26 range from the original to the reduced series more than the broad categorizers.

Twenty-four new subjects were used in this experiment, this time drawn from the fifth grade. The initial series of 20–30 dot clusters was run through randomly eight times, (a total of 88 presentations), and then followed by four presentations of the reduced range of 20–26 dots for a total of 28 presentations. Apart from this difference, the procedure was the same as in Experiment I. The subjects were ranked in the order of their percentage of yes responses to the 20–30 initial range to establish broad, middle, and narrow groups.

Results

The range in breadth of category for Phase I (now 20–30 dots) was from 95.4% to 21.6%. The 27% cut-off point resulted in six subjects being assigned to each of the two extreme groups: the broad and the narrow. The range in breadth of category for the broad group was 95.4%-59%; for the narrow group, 36.4%–21.6%.

This time, as one shifts from Phase I to Phase II, one finds the narrow subjects *decreasing* their calls of 20 in response to the narrowing of the range—all of them doing so with a mean shift in percentage of yes responses of −13.7 (Table 2). The broad group again hardly reacts to change: three of the subjects decrease their proportion of yes responses, one does not change, and two increase the number of yes responses. The mean percentage decrease of yes responses for the broad group is −1.4. The difference, as measured by a *U* test, is significant at the .03 level.

The shift from Phase I to II goes counter to what would be expected by virtue of "ceiling" and "floor" effects, or, indeed, according to the effect of regression. Those who have been calling few of the clusters 20 decrease the number of their calls much more markedly than those who have been calling many clusters 20. While these results provide an effective answer to the query raised in the last section about the possible contribution of artifacts to the results of Experiment I, they also raise some interesting questions of interpretation.

Take first the narrow categorizers. When the range of stimuli to which they are exposed broadens, their categorizing broadens. It is, if you will, now safer to take positive fliers on doubtful stimuli, and more dot clusters are labeled as 20. When the universe of stimuli

TABLE 2
PERCENTAGE OF YES RESPONSES GIVEN BY SUBJECTS IN THE THREE PHASES OF EXPERIMENT II

Group	Phase I (20–30)	Phase I (27–30)	Phase I (20–26)	Phase II (20–26)	Phase III (20–26)
Broad					
1	95.4	90.6	98.2	100.0	100.0
2	80.7	53.1	96.4	85.7	64.3
3	67.0	40.6	82.1	78.6	71.4
4	62.5	31.2	80.4	64.3	57.1
5	62.5	71.9	57.1	57.1	85.7
6	59.0	34.4	73.2	92.9	71.4
M	71.2	53.6	81.2	79.8	75.0
Middle					
1	56.8	37.5	67.9	57.1	50.0
2	55.6	37.5	66.1	78.6	85.7
3	54.4	34.4	66.1	71.4	42.9
4	52.3	28.1	66.1	42.9	42.9
5	50.0	25.0	64.3	42.9	57.1
6	50.0	37.5	57.1	78.6	57.1
7	47.7	12.5	67.9	64.3	42.9
8	46.6	15.6	64.3	57.1	57.1
9	46.6	50.0	44.6	35.7	57.1
10	42.0	28.1	50.0	42.9	57.1
11	40.9	40.6	41.1	57.1	57.1
12	37.5	9.4	53.6	57.1	57.1
M	48.4	29.7	59.1	57.1	55.3
Narrow					
1	36.4	12.5	50.0	28.6	50.0
2	35.2	3.1	53.6	42.9	21.4
3	32.9	34.4	32.1	21.4	21.4
4	32.9	9.4	46.4	42.9	42.9
5	31.8	34.4	30.4	14.3	14.3
6	21.6	12.5	26.8	7.1	14.3
M	31.8	17.7	39.9	26.2	27.4

narrows in range, narrow categorizers show a response of caution, and the bounds of their inclusion class are narrowed. Interestingly enough, both reactions go counter to the "reality" situation. That is to say, in Experiment I the proportion of actual clusters of 20 dots decreases from Phase I to Phase II—yet the yes responses of narrow subjects increases. In Experiment II it is the reverse: the proportion of clusters of 20 dots increases from Phase I to Phase II, but the calls of 20 decrease for narrow subjects. It seems plain then, that while these subjects are responding to change in the environment, they are doing so in terms of *internal* requirements, not *external* ones. What is also interesting is that whereas in Experiment I narrow subjects reverted to narrow categorizing after they had habituated to the extended range, in Experiment II they do *not* change in Phase III after they have narrowed their categorizing in response to a narrowed range of stimuli. There is no change in the percentage of yes responses from Phase II to Phase III (Table 2). In short, a change in categorizing in the preferred direction sticks, a change in the nonpreferred direction (toward greater breadth) habituates out.

Now consider the broad categorizers. In Experiment I we show that they did not respond on the average to an extension of the range of stimuli to which they were exposed in shifting from Phase I to Phase II. But in Phase III, habituation to the extended series, or possibly an insignificant and delayed reaction to the change of stimulation, led to a moderate increase in the breadth of their 20 category. They, too, respond to increased diversity in environmental stimulation, but more slowly and more modestly. In Experiment II, the picture is very similar: as before, and in contrast to narrow categorizers, their response to change is slow and insignificant.

One last check on possible artifacts in Experiment I is provided by the design of Experiment II—a check on regression effects. In Experiment I, the transition from the initial to the extended range occurred after 8 successive presentations of the 7 stimuli of the initial range—a total of 56 presentations. This was followed by Phase II which consisted of a series of 11 stimuli repeated twice—22 presentations. As Phase I in Experiment II consisted of 11 stimuli repeated 8 times, it was possible to ascertain whether any regression effect (which would consist of an increase in the proportion of yes responses for the narrow group and of a decrease for the broad group) had occurred at the point of presentations nearest to the transition point between Phases I and II in Experiment I. This was done by comparing the percentages of yes responses for the first 5 presentations of the series (a total of 55) with the same percentages for the 2 following presentations (a total of 22). There was no evidence of any trend towards increase for the narrow categorizers, or towards decrease for the broad ones.

EXPERIMENT III

This experiment was conducted in order to ascertain whether the results of Experiment I were due to consistent differences between the broad and the narrow categorizers in their preferred mode of dealing with an uncertain

cognitive situation, rather than to a tendency of the narrow group to be more accurate, presumably under any conditions. A subject who is in the group of narrow categorizers in Experiment I because he might be intent to perform as accurately as possible should become a broad categorizer in a situation where the distribution of stimuli is such that most stimuli fall into the inclusion or target category. The converse is, of course, not necessarily true: if a broad categorizer performed as he did because he was bored and was paying very little attention to the task, he might still perform in a similar manner when confronted with any distribution of stimuli—assuming that yes responses are easier for maintaining an attitude of nondiscriminating boredom.

Though we have spoken of "accuracy," there is one general question concerning it that is difficult to answer in these experiments. In most cognitive situations flexibility and avoidance of extreme response tendencies probably lead in the end to most "adaptive" responses—that is, to responses which keep in closest touch with the normally distributed and continuously changing environment. There may be situations, however, in which either a very narrow or a very broad type of response is the most useful. This, as Bruner, Goodnow, and Austin (1956) have shown, is a function of the "payoff matrix" in the situation, of the respective consequences that errors of overinclusion or of overexclusion may have. It is, therefore, impossible to state a priori which of the two extreme forms of categorizing is the "better" or the most useful save in a restricted way applying to specific situations.

The arrangement of stimuli in Experiment I favored the narrow categorizer in the sense that few stimuli were in fact 20s. We decided, therefore, to run another experiment in which a broad style of categorizing would lead to performance corresponding more closely to the actual distribution of stimuli. This time, the great majority of stimuli would belong in the inclusion class and the preferred response distribution of broad categorizers would conform to the distribution of "correct" or inclusion stimuli.

Subjects and Procedure

Experiment III was conducted with 45 subjects who took part in Experiment I plus 4 subjects who were added subsequently. The latter were first subjected to the procedure of Experiment I, and then, a few days later, took part in Experiment III. Stimuli consisted of straight lines drawn on rectangles of white cardboard, 11″ × 14″ in size. The inclinations and positions of lines on the rectangles were varied randomly. There were 56 rectangles: on 48 of them (6 out of 7) lines 5″ long were drawn, one to a rectangle. The remaining 8 rectangles contained lines varying in length from 5¼″ to 5¾″. The subjects, tested individually, were seated at one end of a table, the experimenter at the other, separated by a distance of 3 feet. The lines were shown one by one, the order of presentation so arranged that one line longer than 5″ appeared in each successive series of six 5″ lines, the serial position of the longer line varying at random. Subjects were told to decide whether a line was 5″ long or not. If they thought that it was, they should say "yes"; if they thought not, "no." They were duly informed that some of the lines would be 5″ long, some slightly longer. Two 5″ lines were shown first, and the subjects told what they were. The presentation of the experimental series followed immediately. Time of exposure was not limited, the cardboard rectangle being held up until the subject made his decision.

Scores again consist of the percentages of yes responses given to the total of 56 stimuli presented—a number equal to the number of clusters of dots presented in Phase I of Experiment I. The prediction, of course, is that if the subjects' performance in Experiment I reflected their preferred mode of categorizing, they should perform in a similar manner in Experiment III, despite the radical change in the frequency of stimuli belonging to the inclusion class—that is, broad categorizers should respond more broadly in both experiments.

Results

The range of percentage of yes responses was from 100% to 35.7% with a median of 57.1%. Table 1 sets out the distribution of the subjects according to the percentage of their yes responses.

1. *Comparison of breadth of category in Experiment I and Experiment III.* The subjects were ranked according to their breadth of category in both experiments. A rank correlation was run between their performances in the two experiments. This was found to be of the order of rho = .572, significant at $p < .0005$.

A further assessment of the similarity of performance in the two tasks was made by selecting the subjects who in Experiment I belonged to the broad and narrow groups, and comparing the difference in the performance of these two groups in Experiment III. The cut-off point for the groups in Experiment I was

slightly different from the one used in the previous analysis of results for this experiment, as some subjects dropped out and some were added. Taking ties into account, the nearest point to 27% now included 15 subjects in the broad group, the narrow group consisting as previously of 13 subjects. The difference, as assessed by a *U* test, is highly significant, the confidence level being less than .001. The mean percentage of yes responses in the previously defined broad group is 69.5. The percentage of yes responses in Experiment III of the previously defined narrow group is 51.8.

2. *Comparison of accuracy of performance in Experiment I and Experiment III.* The group of narrow categorizers performed more accurately in Phase I of Experiment I than the group of broad categorizers. Accuracy of performance was assessed by assigning to each subject a score that consisted of the proportion of his number of yes responses assigned by him to clusters containing 20 and 21 dots. The percentage of accurate calls for the narrow group was 45.9, as compared with 33.4 for the broad group, a difference significant at the .001 level.

The same measure of accuracy could not be applied in Experiment III, as almost all subjects gave all correct no responses to the eight lines that were longer than 5″. Therefore, the assessment of accuracy here was made by summating *all* incorrect responses made by a subject: no responses given to 5″ lines, and yes responses given to the longer lines. Of the 25 subjects belonging to the broad and narrow groups in Experiment I whose accuracy of performance in that experiment had been originally compared, 23 were still available in Experiment III. Their accuracy of performance in the new task was compared and the broad subjects were found to have 81.5% of their calls accurate, in contrast to 71.4% for the narrow subjects—a difference significant at the .01 level.

To sum up Experiment III, then, subjects maintain their relative positions with respect to breadth of categorizing when they shift from a situation where few stimuli fit the specifications of an inclusion class to one where many stimuli fit that specification: ones who are narrow in the first situation tend to be narrow in the second, and so, too, for those who are at the broad end of the distribution. One finds, moreover, that whereas in a situation where few stimuli fit the specifications for inclusion in the target class, narrow categorizers are more accurate, whereas when many stimuli fit, the broad categorizer turns out more accurate. Accuracy, in brief, seems more a function of a happy fit between response preference and stimulus properties than of discrimination capacity, much as in an experiment of Tagiuri, Bruner, and Blake (1958), where accuracy of social perception was a systematic resultant of two response patterns that had nothing to do with discriminative finesse. It would seem, then, that *not* only does discriminative capacity not determine categorizing breadth, but that, in a paradoxical way, it is the other way round.

CONCLUSIONS

Several things appear quite plain. To begin with, there appear to be consistencies in breadth of categorizing. We find, moreover, that narrow categorizers tend to be more sensitive to changes in the stimulus environment. They show more alteration in their categorizing breadth in response to change and show it more swiftly—though in a way that is not necessarily "rational." That is to say, when the range of stimuli to which they must respond *broadens*, they broaden their equivalence range for stimuli of a target class, although the specifications of the target class have not changed, and although their response may lead to more errors. It is as if they were "exploring" the new and wider range. Broad categorizers, on the other hand, do not change as much in response to an increase in the range of stimuli to which they must respond. They appear to be "holding on" to old methods in the face of a new situation. When the changed stimulus situation persists and becomes a steady state, the narrow categorizer reverts to his more precise and narrow mode of categorizing and reduces error. The broad categorizer, slower to react in the first place, begins to respond to change only when the change becomes more permanent.

An analogous picture holds when the stimulus world to which broad and narrow categorizers are exposed is constricted and becomes less varied. The narrow categorizer "follows" the change more swiftly by narrowing his categorizing range for the target class. Again even though logically it may lead to more error, for the specification of the target class

has not changed. This time his response seems to be in the direction of closer analysis of a narrower range. The broad categorizer responds with no discernible change, as if he were riding out the narrowing of the stimulus world that runs counter to his preference for breadth rather than changing with it.

It should not be surprising then—although we were taken unawares—that breadth of category is associated with intelligence, the narrow categorizers having a higher average IQ than the broad ones. Interestingly enough, though, shift in categorizing in the face of stimulus change is found to be related to breadth of categorizing but not to intelligence. Regardless of intelligence, the narrow categorizer handles change by shifting, by reacting to contrast; the broad categorizer shifts much less and later. In sum, reaction to change appears to be a strategy of dealing with the consequences of error. The narrow categorizer appears to prefer the risk of *reacting* and possibly being wrong. The broad categorizer prefers the risk of *not reacting* to change and possibly being wrong. Now we are in a position to ask about personality. What personality characteristics produce such preferences?

REFERENCES

ARNHOFF, F. N. Ethocentrism and stimulus generalization. *J. abnorm. soc. Psychol.*, 1956, **53,** 138–139.

ARNHOFF, F. N. Ethnocentrism and stimulus generalization: A replication and further study. *J. abnorm. soc. Psychol.*, 1957, **55,** 393–394.

BRUNER, J. S., GOODNOW, JACQUELINE J., & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18,** 108–143.

GARDNER, R. W. Cognitive styles in categorizing behavior. *J. Pers.*, 1953, **22,** 214–233.

HAMILTON, V. Perceptual and personality dynamics in reactions to ambiguity. *Brit. J. Psychol.*, 1957, **48,** 200–215.

KELLY, T. L. The selection of upper and lower groups for the validation of test items. *J. educ. Psychol.*, 1939, **30,** 17–24.

KLEIN, G. S. Cognitive control and motivation. In G. Lindzey (Ed.), *Assessment of human motives.* New York: Rinehart, 1958. Pp. 87–115.

ROKEACH, M. "Narrow-mindedness" and personality. *J. Pers.*, 1951, **20,** 234–251.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

TAGIURI, R., BRUNER, J. S., & BLAKE, R. R. On the relation between feelings and perception of feelings among members of small groups. In Eleanor E. Maccoby, T. Newcomb, & E. Hartley (Eds.), *Readings in social psychology.* New York: Henry Holt, 1958. Pp. 110–116.

TAJFEL, H. The anchoring effects of value in a scale of judgments. *Brit. J. Psychol.*, 1959, **50,** 294–304.

ZASLOW, R. W. A new approach to the problem of conceptual thinking in schizophrenia. *J. consult. Psychol.*, 1950, **14,** 335–339.

(Received January 20, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 242-249

EXISTENTIAL PSYCHOTHERAPY:

FACT, HOPE, FICTION

JOSEPH LYONS

Veterans Administration Hospital, Lexington, Kentucky

More than 20 years after it was first published, a monumentally difficult, subtle, and as yet untranslated German philosophic work—*Sein und Zeit* by Martin Heidegger—has become the bible of many psychotherapists in this country. Heidegger has nothing to say explicitly about therapy; and even those of his ideas which seem relevant have been accessible to English-speaking readers only in some of his shorter pieces, through expository writings in various textbooks of philosophy and theology, or in one excellent essay by Brock (Heidegger, 1949, pp. 24–131). Therefore Heidegger's influence has had to be exerted indirectly, through one or another of the practicing psychiatrists who claim to speak for him or in his name. In this way, often at one or more removes, a new approach to psychotherapy has recently come into prominence and a new vocabulary has become fasnionable: encounter, authentic and unauthentic, care, being, engagement, finitude, dread, thrownness, and commitment, among many other terms.

Most striking is the fact that although a supposedly novel and radically different approach to the practice of psychotherapy has lately appeared in America—bearing the name of existential analysis as an adaptation of a significant trend in Europe—its guiding ideas are either guesses as to what the founding figures might have advocated or suggestions culled from scattered works which were not meant originally as texts in psychotherapy. As a full-scale exposition of theory and practice, there is next to nothing from any of the major European figures in this movement, with the sole exception of Frankl (1958), who has been careful to identify his own approach as Logotherapy and to insist that the name for this type of existential analysis is not *Daseinsanalysis*, as with Binswanger, but *Existenzanalysis* (1959, p. 157).

Existential analysis appears to be the most important movement in the therapeutic field since the rise of the neo-Freudian schools of Sullivan and Horney. Indeed, in the eyes of its wide range of proponents, the historic mission of the new approach is to go beyond all forms of psychoanalysis both in doctrine and in practice. In a report on the recent International Congress of Psychotherapy at Barcelona, Lefebre (1959) noted that, to judge from the papers at this meeting, the existential analytic approach enables psychotherapists to "rescue psychoanalysis from Freud and the Freudians." Yet in spite of the enthusiastic support it has enlisted, there remains the strong possibility that the new movement may be not yet a fact but only a hope, or that on examination it may be revealed as largely a fiction. It is the purpose of the present paper to review a selection of claims made for existential psychotherapy and to attempt an assessment of its present status and future prospects.

Definition

The first question that may be raised is what is meant by existential analysis or existential psychotherapy. Surprisingly, however, there seems to be little agreement on a definition of the term. Its history may therefore be of interest in pointing to some of the terminological and conceptual confusion in this area. According to Straus (1959), the term which has been translated as Existential Analysis was introduced by Binswanger (1958a) in his "Case of Ellen West." However, he meant it in quite a special sense, and he explicitly distinguished between existential analysis (*Daseinsanalyse*, as he called it) and Heidegger's analytic of existence (*Analytik des Daseins*). "The similarity of the expression," Binswanger noted, "is justified by the fact that the anthropological or existential analysis relies throughout on that structure of existence as being-in-the-world which was first worked out by the analytic of existence" (1958a, p. 271). A great part of Heidegger's *Sein und Zeit* is devoted to an elaboration of the "existentials"[1] which make

[1] This term (*Existentialen*), referring to aspects of the structure of *Dasein*, is used in place of "categories"; for technical reasons the preference is not arbitrary.

up the structure of *Dasein*; and it was this elaboration, the term for which is usually translated as analytic of existence, which provided the starting point for Binswanger's (1957) historic series of clinical studies. However, the clinical work of the pupil differs quite deliberately from the ontological investigations of the teacher. Binswanger's approach rests upon and arises from Heidegger's studies and often makes use of somewhat similar terminology, but it is meant essentially as a translation into concrete instances of the guiding conceptions which were originally worked out for the metafactual realm of *Dasein*.

If only this were the case, it would be stretching the meaning of the term to reapply Heidegger's teachings under the name of existential analysis as though they constituted a form of therapeutic doctrine. But to complicate the matter further and to make even less seemly the present trend toward an indiscriminate use of Heidegger's conceptions, Binswanger's (1958b) term has itself undergone a subtle alteration. At no point did the latter represent it as a method or approach in therapy; rather, he offered it as "an anthropological type of scientific investigation" (p. 191). As a practicing clinician, he is well aware of the essential difference in attitude between treatment in therapy and investigation in science. At best, he maintains, the success of his method of investigation might aid the clinician to "comply with a therapeutic demand," but in this respect existential analysis is no more a method of therapy than is an investigational technique such as the Rorschach. More recently, in his paper for the Barcelona congress, Binswanger re-emphasized this point. Existential analysis, he said, refers to a scientific, not a therapeutic endeavor; psychoanalysis is still indispensable.

These statements may be contrasted with the varied definitions for existential analysis that have been offered since the approach has become widespread. It has been defined as "the application of existentialism to psychotherapy" and as "an exploration of the ontological ground of man" (Van Dusen, 1959a, pp. 101, 104). According to one reviewer (Ellenberger, 1958), it is "simply the application of certain existentialist concepts to psychotherapy, without regard to phenomenology or existential analysis" (p. 119); yet another writer (May, 1958a) defines the existential analyst as a phenomenological and existential psychiatrist or psychologist. Finally, existential analysis has been called "a clinical practice, outwardly similar to other forms of psychoanalysis" (Van Dusen, 1957, p. 310). The lack of agreement here suggests a serious confusion at just the point where there ought to be the most clarity.

What, in actual fact, does existential analysis as described by its proponents consist of? Five aspects may be discussed in terms of the extent to which each appears to be a sound contribution to clinical theory: a description based upon a certain view of modern culture, a set of working procedures, a new terminology for current therapeutic concepts and practices, an attack on some excesses and limitations of orthodox Freudian theory, and an attempt to popularize the thinking of Heidegger.

A Description Based upon a View of Culture

In one of the early papers in English on this subject, Edith Weigert (1949) noted that existentialism does not lend itself to a concern for individual psychopathology but only to a description of "certain basic irreducible dialectic trends like One and Self, Care and Love" (p. 409). "Even if the existentialist may not have much to contribute to a greater understanding of individual psychopathology or therapy, he affords us a profound insight into the pathology of the culture as a whole" (p. 412). This line of reasoning, to the effect that a method for understanding what is sometimes called the sickness of our age is helpful to the therapist and perhaps even to his patient, forms the basis for the essays on modern man with which existential analysts frequently preface their clinical studies. The lengthy historical survey by May (1958a) is a case in point. Writers in this area, often showing more concern for man in general than for men in particular, show a strong tendency to turn their study of persons into a self-conscious sociology.

One wonders, however, how relevant the theme of cultural history is to the practice of psychotherapy. It is undeniable that at every point in his therapeutic progress the patient is at a moment in human history as well as at a moment in his own history. But are these

two moments equally pertinent to the process of therapy? The patient's insight, "I am estranged from my own true self," may not be the same kind of datum as the insight—on the part of the therapist, or the patient in his own living room, or any other adult—that modern man in general is estranged. Indeed, it is difficult to see how attaining the latter insight could help one to achieve the former, especially since it is also probable that it would act as an obstacle rather than an aid.

A SET OF PROCEDURES

In the flood of recent commentaries on existential analysis it is not easy to find a thorough, precise statement of its procedures, particularly as they may differ from current practices among competent therapists of all persuasions. May (1958a) refers in general terms to a "variability" of techniques, the specific ones depending on the patient and the phase of treatment; but he also quotes a personal communication from Roland Kuhn, a Swiss psychiatrist who has worked closely with Binswanger: "Since existential analysis is a relatively new discipline, it has not yet had time to work out its therapeutic application in detail" (p. 76). Van Dusen (1959a) is somewhat more specific, though he too begins with the general statement that "other than some central elements, the existential analysts have not yet developed a dogma of practice" (p. 108). He sees the following as the central elements of practice: a variety of techniques and approaches, a centering on the critical here and now, an emphasis on the patient's capacity to choose rather than on directive methods, an avoidance of categories and the historical past (1957, p. 318), and "a *Weltanschauung* which sees a being here and now struggling to manifest more perfectly" (1959a, p. 108). The essential characteristic, however, "lie(s) in the way the analyst understands the process" (1957, p. 319). Boss (1958), who is one of the European leaders in the new movement, has stated frankly that "existential analysis as such is indifferent both to psychotherapeutic techniques and to practical consequences and aims" (p. 119). Other writers, however, express a greater optimism, although they are seldom able to be more specific. Ellenberger (1958) notes hopefully that "phenomenology opens the path to a new type of psychotherapy which is still in its early stages of development" (p. 113), and Straus (1959) offers an optimistic summary in regard to therapeutic practice: "Although *Daseinanalyse* can be applied to psychiatric cases, at its roots it is not a specific psychiatric method; still less is it primarily a psychotherapeutic technique. However, the concept of being-in-the-world, as the basic structure of *Dasein*, the distinction of authentic and unauthentic modes of existence, the emphasis placed on finiteness, Nothingness, and on the relation to possibilities and to the future—all these lead to definite psychotherapeutic consequences" (p. 161). In any case, there remains the difficult question of whether, if the practice of existential analysis does not differ from other therapeutic methods, we are not all really existentialists without knowing it; or if we are not, whether the existential analysts are implying that methods and procedures are in the long run less important than the therapist's theoretical constructions and modes of understanding.

It may be noted that there is a discrepancy between two of the claims made for existential analysis. On the one hand, its expositors insist on their fundamental respect for the concrete phenomenological datum—although they usually use the term "phenomenological" in the limited and nontechnical sense of "subjective"—which is obtainable, in the view of some, only by an act of empathic apprehension of the structure of the patient's private world. Yet on the other hand, they emphasize their concern for general and historically conditioned human attitudes and for constructions which are intellectually satisfying to the therapist. The discrepant claims usually seem to be resolved in favor of a guarantee of growth on the part of the therapist, whether or not there is also change in the patient. The nature of this resolution—best exemplified in Binswanger's (1958a) "Ellen West," which has perhaps served as the model—may throw some light on the fact that on close examination existential analysis turns out to have very little to offer in regard to the practicalities of everyday clinical work. It is of some interest that Frankl (1958), who has publicly dissociated himself from the main stream of existential analysis, is the only one of the European psychotherapists who has made

specific contributions to therapeutic practice, for example, his technique of the "paradoxical intention." At the same time, however, he is careful to avoid any claims to establishing a new school or method of therapy: "Logotherapy is no substitute for psychotherapy but its complement" (p. 36).

TERMINOLOGY FOR THERAPEUTIC CONCEPTS AND PRACTICES

When the conceptions and practices offered as existential analysis are summed up and critically examined, they turn out to bear a close resemblance to many which have been adopted independently by psychotherapists trained in widely varying disciplines and approaches. Consider, for example, this description by Sonneman (1958) of the essence of the therapy process: "The daseinanalytic therapist-patient exchanges are therefore quite frontal discussions on a basis of absolute and equal freedom—free from any and even the most hidden implications of a mere status-derived or situation-based authority of the therapist. They are shared engagements of both partners at once, and what the two engage in is never the psyche of the patient in a fictitious isolation but always his world, always that which matters to the patient, and exactly on the terms on which it does" (p. 145). It is true that the practice of therapy does not always achieve this laudable goal, for reasons attributable many times to the narrowness or incompetence of the therapist himself, and it is welcome indeed to have the case for an optimal relationship stated so movingly. But it is equally true that no competent therapist would deny these words, or even fail to understand them immediately and be able to apply them to his own practice, to his successes as well as his failures. The basic question, which is usually bypassed, is whether the admitted validity of Sonneman's ideas bears any legitimate relation to Binswanger rather than, say, to Sullivan, to the best of Heidegger rather than the best of Freud or Adler.[2]

Still clearer examples may be found in other writers. May (1958a), after stating that "the achieving of the sense of being is a goal of all therapy" (p. 44), spells out his prescription, which is essentially a recommendation for genuineness on the part of the patient. Similarly, Hora (1959b) substitutes a new term "unauthenticity," for the old term "duplicity" and defines it as what occurs when the patients gratify their own "personal needs for tension-reduction even though they appear to be talking to each other and seemingly are interested in each other" (p. 85). He makes an excellent plea for such established goals in therapy as sound values, insight, self-awareness, and the constructive use of anxiety, goals with which none would quarrel. He insists, however, on presenting them in a new wrapping. But in so doing, does he not run the risk of contradicting the very goals that he advocates for his patients?

ATTACK ON ORTHODOX FREUDIAN THEORY

Something like the existential analytic movement was, without doubt, long overdue in this country. It had already been an important influence in European clinical practice for more than a decade. Although a number of important movements had developed before this to challenge the orthodox Freudian position, they all shared the weakness that they were in significant measure reactions against Freud and therefore at their origins presumably joined in a common system. In order to avoid this dilemma, an approach was needed which started from completely different sources. A psychotherapy based on existentialist tenets seemed to satisfy the requirement and perhaps for this reason was quickly seized upon by those who saw psychoanalysis, in any of its forms, as too restricting because it was bound to outmoded conceptions which told only part of the story about the patient in midcentury. An admirable summary of the necessity for this development is given by Franz Alexander (1959): "One must welcome the existentialist emphasis as a counterbalance to the prevailing trend of psychoanalytic practices and teaching—namely, the trend toward losing sight of the gap between theoretical generalizations and the individual patient, whose uniqueness of necessity requires a flexible application of general principles,

[2] There is no intent here to dismiss some worthy contributions by the method of discovering similar statements in earlier writings by others. The issue is not who deserves credit for saying it first but rather who should be credited at all.

toward replacing the individual approach to each specific person with rigid technical rules. This trend eventually leads to a deification of technique as an aim in itself, as the essence of psychoanalysis instead of a servant of its goals" (p. 94).

The essence of the existential analytic quarrel with Freud, as formulated by Boss (1958), is that the new approach states a human norm which differs from that of psychoanalysis. His own goal, he says, is that his patients "see themselves, so to speak, as a light emanating from the mystery of existence, in whose rays all things and fellow-beings are allowed to appear and to develop according to their own nature" (p. 119). In, thus, repeating the ideas and even some of the phrasing of Heidegger, Boss states clearly a conception of cure, of health, and of psychic functioning which bears little resemblance to the nineteenth century associationism of Freud's theoretical system. Similarly, Binswanger (1958a) devotes an entire section of his longest case study, "Ellen West," to a comparison of his own approach with that of classical psychoanalysis, concluding that the latter is a "one-sided distortion of the human image in the frame of a scientific theory of man" (p. 328). Whether the existential analytic approach will be able to develop its own systematic psychological theory to substitute for the psychology so elaborately worked out by Freud is a question as yet unanswered. Thus far, there is very little evidence in either European or American psychology of such a theory in the making,[3] and the only systematic contributions of merit to have reached a large audience, in the works of Sartre (1948a, 1948b, 1953), have been conspicuously ignored by spokesmen of the existential analytic persuasion.

In the light of the deliberate attempts within existential analysis to break through the limitations of psychoanalytic theory, it is of particular interest that adherents of the new approach have been unable to come to a satisfactory position in regard to Freud's clinical contributions. Lefebre's (1959) comments concerning attempts at the Barcelona congress to rescue psychoanalysis from Freud and the Freudians seem to imply a general agreement among existential analysts that the psychoanalytic practice of therapy is worth taking over into a new system. But Boss (1958), in particular, has some difficulty on this score. After noting that "practical indications" in therapy should usually be consistent with Freudian teachings, he adds that Freud "intuitively" developed technical rules which have a "deeper . . . existential meaning." In regard to his own technical rules, Freud was "intuitively" an existential analyst; only in his theory did he go astray. It would appear from this passage that Boss is trying to do more than save psychoanalysis from Freud; he is trying to save Freud from himself. As a former pupil of Binswanger, who in turn was a lifelong personal friend of Freud, and now as an intimate of Heidegger, Boss expresses in its clearest form that uncomfortable straddling of specific issues which besets the new existential analytic movement.

POPULARIZING THE THINKING OF HEIDEGGER

The fountainhead of such theory of existential therapy as exists today is to be found in Heidegger's writings, although in their concern for the historical setting its expounders also trace the theory back to St. Augustine, Pascal, and Kierkegaard, or link it to more modern thinkers such as Buber, Berdyaev, and even the Zen masters. The immediate reason for the predominant influence of Heidegger is probably not his ideas as such but the fact that both of the important recent influences on the new movement, Binswanger and Boss, have been his disciples. Just about all the recent writings in this country on existentialist conceptions as applied to therapy are based on Binswanger and Boss.[4] Thus, Heidegger has been twice interpreted, and the consequences are far reaching for the ideas under discussion here.

The most striking discrepancy in the contemporary development of a school of existentialist thinking is that Heidegger should have

[3] Mention should, however, be made of the hopeful indications one sees in the writings of the Adlerian group in this country, in the clinical and methodological contributions of Erwin Straus, and in some of the studies by von Gebsattel, Minkowski, and Sarro in Europe.

[4] One exception to this may be Rollo May, who has also been influenced by the thinking of the theologian Paul Tillich.

become the thinker to be followed by psychotherapists. Binswanger (1958a) once pointed to the paradox that Freud developed his image of man from experience with his neurotic patients, thereby completely ignoring many exemplary aspects of his own life pattern. In direct contrast with this, the followers of Heidegger have constructed an optimistic and ennobling scheme on the basis of his philosophy, apparently ignoring both his harsh ethic and the more questionable aspects of his professional career. It is well known that his role, particularly in relation to his teacher and friend, Husserl, left much to be desired at a time of national crisis in Germany when Hitler came to power, and that in his pattern of living since the end of the war he has exemplified the lonely, isolated thinker who is quite removed from those concerns which form the substance of therapeutic problems and practices. More important, his views on the fundamental aims of human striving are difficult to reconcile with familiar goals in therapy. As Weigert (1949) points out, in the end he offers no counsel more noble than to advocate the heroic acceptance of "relative freedom of decision" against the ground of *Sorge*[5] in an indifferent universe. She concludes that "a psychiatrist may feel that there is something unconvincing in this heroic power operation, the isolated transcendence of fundamental human helplessness" (p. 409).

The attempt to make of these views a palatable concoction for the therapist and his patients has resulted in some odd twists which range from errors of emphasis to serious distortions. To take one example, in a discussion of Heidegger's concept of *thrownness*, Hora (1959b) considers man's being "thrown" into the world and having to fit himself into it, his consequent "existential dilemma," and finally the possible ways he might find to extricate himself from the dilemma. There is some practical value in stating the matter in this way, but it misses a fundamental conception, one which is expressed by Sartre in terms of man's being thrust into decision and by Kierkegaard in terms of the ineluctable terror of freedom, to the effect that thrownness is beyond the person's apprehension or use. One's existence is constituted of it; one can only fashion projects in its light. A discussion of thrownness which poses it merely as a problem, structurally similar to a neurosis—that is, as a distressing affection which with effort and skilled help may be overcome—rests on the untenable assumption that it can somehow be encompassed by man's knowledge and earnest endeavors. Indeed, not only is it not reducible to the level of "getting along"; there may ultimately be no getting along, not in this stern view of man and his fate.[6]

Other aspects of the twice-removed interpretation of Heidegger contain misconceptions which if logically pursued would deny his aim while professing to expound his views. As is now well known, Heidegger's self-assigned task in *Sein und Zeit* was "the quest for Being," to revive through ontology the ancient problem of Being which had almost disappeared as a matter of concern in contemporary European philosophy. An abstruse and fundamental study, ontology is concerned, not with entities, such as persons or objects or living things, which manifest Being, but rather with that which they manifest. However one wishes to phrase it—as inner structure, guiding idea, or underlying phenomenon—Being as a subject of study is not resolvable into what can be classified or described as an entity, but must be understood as that which infuses the entity and finally makes classification and describing possible.

In the course of his analysis, Heidegger maintains one distinction without which his quest could not proceed—that between the ontological, which concerns Being, and what he terms the ontic, which refers to that set of given and classifiable objects known as particular beings. The distinction is not identical with that between the abstract and the real, for in that case it would point to nothing but a problem in logic; nor is it between the ideal

[5] *Sorge* (literally "anxious care") is a basic term in Heidegger's work. It is the name for the whole structure of *Dasein*, ontologically, that is, for the Being of *Dasein* itself.

[6] It is perhaps less than accurate to refer to Heidegger's view in this connection as stern, for as an ontological study it assumes no stand in regard to values. Ontology as a discipline proposes to explore the very ground for human existence; it must therefore precede rather than follow on such aspects of human living as values. In softening the implication of Heidegger's teachings by adding the mode of Love to the mode of Care, Binswanger would seem to have fallen into the error of reversing this logical sequence of ontology and values.

and the material of metaphysics. Rather, it refers on the one hand to Being as fundamentally manifest in facts and entities present to us and on the other to the concrete and factual, through which we have our dealings with Being. This is a distinction which is usually clear and justifiable, except in the special case of man when he is made an object of study. For a particular, individual man is an ontic phenomenon, yet his particularity is never quite that of the single entity. He is unique as well as individual, creator as well as victim of his own special fate, and he demonstrates his nature in the very process of being aware of this. Within him and through him there abides a special relation to Being. The totally new conception which Heidegger now develops to refer to the phenomenon which sustains this relation is *Dasein*. Although he appears to be talking about human beings and their problems, in his use of this term he does not mean either man, or person, or particular being, or (as the word *Dasein* might ordinarily be translated) existence in the sense of mode of living.[7] In what amounts to his prolegomenon to a science of psychology, he presents his analytic of *Dasein*, particularly in respect to its basic structure of being-in-the-world. The term world is here understood, as one commentator (Macquarrie, 1955) puts it, as "an instrumental system founded in man's practical concern" (p. 49), as a nexus of relations rather than a locus or a set of activities. In all of this, it should be noted, man as an observable entity who is studied by the various sciences does not appear, nor are concrete human circumstances the substance of the author's concern.

It would seem unlikely that the analytic of *Dasein* could be applied directly to specific instances such as the symptoms of particular patients. Yet this bold move is what Binswanger (1958b) attempted in his development of the existential analytic school of thought, or *Daseinsanalyse*. However, he has not been unaware of the significance of Heidegger's distinction between the ontic and the ontological, as his disclaimer clearly shows: "Existential analysis does not propose an ontological thesis about an essential condition determining existence, but makes ontic statements—that is, statements of factual findings about actually appearing forms and configurations of existence" (p. 192). Recently, in fact, he stated that his entire attempt at an ontic "translation" may have been an error but that it was a productive error.

In the course of reinterpreting Binswanger's interpretation, however, recent writers on existential analysis appear to have overlooked this distinction between ontic and ontological, changed the meaning of ontology almost beyond recognition, and engaged in a wholesale application of phenomenological terms to clinical concepts which are only verbally similar. Examples are numerous in the recent journal literature: "Being-there"[8] refers to "specific ways of experiencing life and specific ways of dealing and communicating with the environment" (Hora, 1959b, p. 83); and again, " . . . mode of being-in-the-world is the individual's unique way of encountering and dealing with the world" (Van Dusen, 1959a, p. 105). The "existentials" of Heidegger are equivalent to "inborn qualities" of the person, something like instincts (Weigert, 1949). The "ontological basis" of the psychological insights achieved by existential analysis is "the study of man as the being who has these particular problems" (May, 1958b, p. 36). "Self-transcendence"[9] means that one transcends one's immediate and selfish interests (Hora, 1959b). "In everyday terms, being embraces non-being in the sense that we can be aware of death . . . can by self-awareness encompass death" (May, 1958a, p. 49). A style of life, in Adler's sense, and the striving for superiority as well, are "ontologically grounded," for Adler isolated "ontologically fundamental aspects of man . . . showing their dynamics" (Van Dusen, 1959a, p. 105). Existential psychotherapy "unites science and ontology" (May, 1958b, p. 36). "Man's mode of being-in-this-world depends very much on his epistemological point of view, that is . . . the nature of his knowledge . . . " (Hora, 1959a, p. 169). Being is synonymous with "what exists," and ontology "when applied in psychology . . . is

[7] The *hypostasis* of Thomistic metaphysics, which is the term of highest dignity for the subsistence of a person of the Trinity, furnishes a precedent sufficiently similar to be of more than passing interest to the student of the history of ideas.

[8] This hyphenated term is a literal rendering of the German *Da-sein*.

[9] Temporality is of central importance in Heidegger's thinking, so much so that he has coined the term *Zeitigung* to express the "engendering" of time through

the study of man's being or nature" (Van Dusen, 1959b, p. 143).

These statements, which constitute a fair sampling of what has been offered in the recent literature as the philosophic basis for existential analysis, have no legitimate connection with Heidegger and manifest only the barest resemblance to Binswanger. As ideas, they may be useful in themselves. For the demonstration of this, one may only wait and see. But to claim more for them at the moment would seem to bespeak at best a neglect of the sources from which they are supposed to have been learned and at worst an ill-founded attempt to make use of the prestige of those sources.

Clearly, existentialism as a movement in the history of ideas has profound and perhaps enduring contributions to make to every significant human endeavor. Some aspects of the vision which inspires its protagonists may be found, for example, in Lefebre's (1959) conclusion: "... the conception of man as one who seeks for his own truth and who falls ill if he cannot bring his truth into living reality is what distinguishes existential analysis from other psychotherapeutic schools of thought" (p. 118). Binswanger's conception of the "vocation" of the therapist, as described by Weigert (1949), also reaches this level: "... open to visions of togetherness, totality, to existence as love, to the healing potentialities of the creative Eros" (p. 408). But there is still the possibility that this inspired and historically necessary expansion of the healing mission of the therapist will be dissipated in hollow claims. Existential psychotherapy today is more gaudy fiction than fact; yet it remains a hope.

REFERENCES

ALEXANDER, F. Impressions from the Fourth International Congress of Psychotherapy. *Psychiatry*, 1959, **22**, 89–95.

BINSWANGER, L. *Schizophrenie*. Pfullingen: Neske, 1957.

BINSWANGER, L. The case of Ellen West. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence*. New York: Basic Books, 1958. Pp. 237–364. (a)

BINSWANGER, L. The existential analysis school of thought. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence*. New York: Basic Books, 1958. Pp. 191–213. (b)

BOSS, M. *The analysis of dreams*. New York: Philosophical Library, 1958.

ELLENBERGER, H. F. A clinical introduction to psychiatric phenomenology and existential analysis. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence*. New York: Basic Books, 1958. Pp. 92–124.

FRANKL, V. E. On logotherapy and existential analysis. *Amer. J. Psychoanal.*, 1958, **18**, 28–37.

FRANKL, V. E. The spiritual dimension in existential analysis and logotherapy. *J. indiv. Psychol.*, 1959, **15**, 157–165.

HEIDEGGER, M. *Existence and being*. Chicago: Regnery, 1949.

HORA, T. Epistemological aspects of existence and psychotherapy. *J. indiv. Psychol.*, 1959, **15**, 166–173. (a)

HORA, T. Existential group psychotherapy. *Amer. J. Psychother.*, 1959, **13**, 83–92. (b)

LEFEBRE, L. Report on the Fourth International Congress of Psychotherapy. *Amer. J. Psychother.*, 1959, **13**, 111–120.

MACQUARRIE, J. *An existentialist theology: A comparison of Heidegger and Bultmann*. New York: Macmillan, 1955.

MAY, R. Contributions of existential psychotherapy. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence*. New York: Basic Books, 1958. Pp. 37–91. (a)

MAY R. The origins and significance of the existential movement in psychology. In R. May, E. Angel, & H. F. Ellenberger (Eds.), *Existence*. New York: Basic Books, 1958. Pp. 3–36. (b)

SARTRE, J-P. *The emotions: Outline of a theory*. New York: Philosophical Library, 1948. (a)

SARTRE, J-P. *The psychology of imagination*. New York: Philosophical Library, 1948. (b)

SARTRE, J-P. *Existential psychoanalysis*. New York: Philosophical Library, 1953.

SONNEMAN, U. The human sciences and spontaneity: Outline of a revolution. *Amer. J. Psychoanal.*, 1958, **18**, 138–148.

STRAUS, E. W. The Fourth International Congress of Psychotherapy. *Psychosom. Med.*, 1959, **21**, 158–164.

VAN DUSEN, W. The theory and practice of existential analysis. *Amer. J. Psychother.*, 1957, **11**, 310–322.

VAN DUSEN, W. Adler and existence analysis. *J. indiv. Psychol.*, 1959, **15**, 100–111. (a)

VAN DUSEN, W. The ontology of Adlerian psychodynamics. *J. indiv. Psychol.*, 1959, **15**, 143–156. (b)

WEIGERT, EDITH. Existentialism and its relations to psychotherapy. *Psychiatry*, 1949, **12**, 399–412.

(Received January 25, 1960)

which Dasein is always in advance of itself and therefore always what it is not. "Self-transcendence" is meant to point toward this concept.

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 250-256

GENERALIZATION OF INTERPERSONAL SIMILARITY[1]

EZRA STOTLAND
University of Washington
ALVIN ZANDER
University of Michigan
AND THOMAS NATSOULAS
Wesleyan University

One of the most commonly observed and widely studied phenomena in interpersonal relations is that the degree of similarity among persons is a powerful factor in their reactions to one another (Asch, 1952; Back, 1951; Raven, 1959; Schacter, 1951; Sherif, 1936). The present paper is concerned with the development of similarity among persons when one person models himself after another. Our general hypothesis is that the perception of similarity among persons in limited respects leads to the perception of similarity among them in other respects. This tendency to generalize similarities is assumed to be one form of the phenomenon often described as identification.

In keeping with the cognitive emphasis of the study, we employ the self-concept as a central explanatory concept. A person's self-concept is assumed to include a number of attributes. The individual also develops concepts of other persons. He may then perceive that one or more of the attributes of his self-concept is similar to one or more attributes of his concept of a particular other person, termed the model. This similarity is designated as a first similar attribute (FSA). Our key proposition holds that the presence of an FSA creates a tendency for the person to conceive of himself as similar to the model on other attributes, provided that there is no clear evidence to the contrary. These additional similar attributes are termed derived similar attributes (DSAs). It is further assumed that, the more FSAs there are between a person and a model, the more DSAs will develop.

The process of generalizing similarities is assumed to take either of two forms, introjection or projection. Introjection is the cognitive process whereby an attribute previously present only in a person's concept of the model now becomes an attribute in his self-concept. Projection is the process whereby an individual changes his concept of a model to include attributes previously present only in his self-concept. These two processes are assumed to be types of identification.

The study concerns two hypotheses concerning introjection and projection:

Hypothesis 1: The greater the number of FSAs in a person's perceptions of himself and another individual, the greater is the tendency for the perceiver to introject DSAs into his concept of himself.

Hypothesis 2: The greater the number of FSAs in a person's perceptions of himself and another individual, the greater is the tendency for the perceiver to project DSAs into his concept of the model.

The validity of one additional assumption is examined: that each DSA introjected into the self-concept can serve as an FSA insofar as future introjections and projections are concerned, thus, increasing the probability that additional DSAs will be introjected. The introjective process, in short, can be cumulative with each DSA making it more probable that there will be further DSAs.

Method

Each subject was provided an opportunity to discover that she had attributes in herself that were similar in some degree to attributes in another. She was next made aware of additional attributes of the model in order to see if she would introject these further attributes (DSAs) into her concept of herself. Then she was made aware of other attributes in her concept of herself and given the opportunity to project these into her concept of the model. Finally, the subject was queried in respect to matters relevant to introjection and projection.

The subjects were 70 undergraduate women recruited from elementary biology courses at the Uni-

[1] This research was financed by Grant-in-Aid M-701 (C) from the National Institute of Mental Health, United States Public Health Service to the Research Center for Group Dynamics, University of Michigan, Alvin Zander, principal investigator.

versity of Michigan on the pretext of an experiment on the nature and causes of personal tastes and preferences.

Procedure. Subjects were brought into the laboratory one by one so as to put each in a private cubicle without allowing her to know who else was participating in the study. Subjects were provided with earphones and a microphone "so they could hear the experimenter and the other subjects, and talk to these persons." In reality, however, everything the subjects heard through the earphones had been previously transcribed and the microphones were dead.

Each subject was informed that she would work with two other persons. The other two were labeled A and B and the subject was told that she was Person C. She was to respond in her microphone, she was told, in answer to questions asked by the experimenter and to write her responses, and those of the others subjects, on forms prepared for this purpose. Because comments by Models A and B were on a tape, any number of subjects could be used, each one responding as though she was a member of a three-person group.

After the subjects were seated, the experimenter began playing the transcription. They heard the experimenter's voice explaining the purposes and procedure.

Phase 1. In order to create the required experimental conditions each subject had first to be made to sense that she had attributes in her concept of herself that were similar to attributes in her concept of a model. Each subject was therefore led to believe that her preferences among seven brief musical compositions were highly similar to the choices of one of the models, and highly dissimilar to the choices of the other model.[2] The subjects heard the experimenter ask them to decide which of two melodies they preferred in each of seven pairs. They listened to each melody twice, made their choice, and recorded it in writing before they heard the other participants' voices reporting their preferences. Periods of silence at appropriate points in the tape made it possible for the subjects to state their choices into the microphone when their turns arrived. The choices made by the Models A and B were also recorded in writing by the subjects, to make them aware of their own musical preference in relation to the preferences of the other two persons.

To ensure that the subjects would be likely to observe by the end of Phase 1 that they were more similar to one of the models than to the other, melodies unfamiliar to all subjects were composed for the purpose. All tunes were pleasant and fairly simple in form. By the use of pretests, seven pairs of melodies were created in which one tune in each pair was very likely to be preferable over the other. One of the models consistently stated preferences from among the more popular tunes while the other spoke for the less commonly liked ones.

In order to rule out preference for one voice over the other, Model A gave the more popular choices for half the subjects and Model B did this for the other half. To avoid suspicion, A and B gave the same preferences on two of the seven pairs.

Phase 2. The purpose of the next step was to make it possible for the subjects now to introject other attributes of a model into their own concepts of themselves. Nonsense syllables with low association value (Glaze, 1928) were used in this instance, to avoid strong preferences based on association. The task was similar to that for the melodies; namely, that all were to state which syllable they preferred in 10 pairs.

The procedure was different from the one used with the melodies in one important respect. Each pair of syllables was typed on a different card and each subject was instructed not to look at the next pair of syllables until the experimenter asked her what her preference was. Since all subjects were in Position C, they heard the choices of A and B, stated as "first syllable" or "second syllable," and recorded these on the appropriate form, before they saw the pair of syllables to which these preferences referred. This procedure was justified to the subjects by explaining that since their first impressions were wanted they were not to know the nature of the syllables until the last possible moment before making their choices.

Each subject, thus, came to know the attributes of A and B without having the attribute present in herself. When it was the subject's turn to respond, she turned the card, read the syllables, and made her own choice. In making this choice she could introject an attribute from the model into her concept of herself.

Each subject had to be made to feel free to make her choices without any concern about how the other would react to them, so that she might not hesitate about disagreeing consistently with one of the models if she wished to do so. This problem was not present in the case of the preferences for music since each subject made her private selection before she heard the choices of the others. In order to avoid possible concern over the model's reactions, the experimenter made a fictitious discovery as this phase of the experiment began, that C's microphone had gone dead. She would still be able to hear A and B, but the latter two would not be able to hear her. After some apparent hesitation, the experimenter "decided" to go ahead, cautioning C to make a careful record of her own choice when her turn came since the others would not be able to do so for her. Each C, then, made her choice after knowing what the choices of the other participants were, but made it in private and with the knowledge that the other persons would not know which syllable she had selected.

Phase 3. The objective for Phase 3 of the experiment was to make it possible for each subject to project attributes of herself into her concept of the model. It was assumed that subjects would do this more often to a person with whom they had many similar attributes than to a person with whom they had few similarities. The subjects therefore had to be made aware of some attribute in themselves that had not previously been present in their concepts of the model, and to provide the possibility for them to assign it to a model.

Each subject was told that she was to pick from pairs of girls' names, recited aloud by the experimenter, the one she liked best. This step was intended to make her aware of a particular attribute in her concept of herself. Next, she was to predict which names A and B would pick, permitting her to project her own preferences onto the models. The experimenter read each

[2] The authors wish to express their appreciation to Edwin J. Thomas and Albert Geibler for their assistance in playing and preparing the music.

pair of names twice and allowed a 30-second period for the choice and predictions. In all cases these choices were written and were not "announced" on the communication system by the subjects.

Phase 4. At this point questionnaires were distributed. The subjects were queried about their strength of preference for the various materials on which they had made their choices, their judgment about which pair of subjects were most similar and least similar, and their perceptions of themselves and the other participants.

After the experiment was over, the subjects were given a full explanation and all their questions were answered.

RESULTS

Main Hypotheses

According to the first hypothesis, the greater the number of FSAs, the greater the number of DSAs introjected. Since there were two potential models for each subject, the number of agreements with each on preferences for melodies was counted. Then the smaller of these figures was subtracted from the larger. This difference score was the measure of the FSAs with the model with whom the subject agreed more. Since the models stated the same preference on two of the seven pairs, the difference scores could be 5, 3, or 1. To obtain the DSAs, the number of agreements with each model on nonsense syllables was computed, and the smaller subtracted from the larger, obtaining a difference score of 8, 6, 4, or 0. A positive DSA meant more agreements with the model who had been agreed with more on melodies; a negative DSA meant more agreements on nonsense syllables with the other model.

To test the first hypothesis, a tau correlation was calculated between the FSAs and the DSAs (considering sign). The correlation (.16) was in the predicted direction, but not significantly so ($p = .09$). Perhaps this resulted from the narrow range of scores on FSAs.

Another test of the hypothesis was performed by dividing the subjects into three categories. The first, called introjectors, were those whose preferences on nonsense syllables were in five or more instances the same as the ones made by the person with whom they had agreed most often in musical choices. The second category, nonintrojectors, contained those whose choices of nonsense syllables agreed with one model four times and with the other model four times. The third category, negative introjectors, included those whose preferences on the nonsense syllables were in five or more instances different from the ones chosen by the person with whom they had agreed most often when making choices of preferred melodies. Table 1 shows that more persons were in the first category of introjectors than would be expected by chance, thus, supporting Hypothesis 1.

A further consideration of the results relevant to Hypothesis 1 concerns an unexpected manifestation of social support for the FSAs. Since both models made the same choices on two of the seven pairs of melodies, any subject could observe that her musical preferences were identical with those of both the models on 2, 1, or 0 occasions. The subjects were divided into those whose musical preferences corresponded to the choices of both models on both possible occasions and all others, and the number of introjectors, nonintrojectors, and negative introjectors in each of these two groups was computed. In Table 2, although the interaction χ^2 is not significant, the data

TABLE 1
NUMBER OF PERSONS WHO DEMONSTRATED DIFFERENT DEGREES OF INTROJECTION

Introjectors	Non-introjectors	Negative Introjectors	N	χ^2	df	p
36 (25.07)[a]	10 (18.86)	23 (25.07)	69	9.09	2	.02

[a] Figures in parentheses indicate expected frequencies. For nonintrojectors this was the proportion of subjects expected to have similar choices four times with one model and four times with the other out of a total of eight choices. The other two expected frequencies: remaining proportion of subjects, halved, and multiplied by the total N. Subsequent tables have expected frequencies calculated in the same way except when stated otherwise.

TABLE 2
EFFECT OF UNANIMITY AMONG MUSICAL PREFERENCES OF SUBJECTS AND MODELS ON NUMBER OF PERSONS WHO INTROJECTED

Frequency of Unanimity on Two Critical Pairs of Tunes	Introjectors	Non-introjectors	Negative Introjectors	χ^2	df	p
2	10 (5.09)	1 (3.82)	3 (5.09)	7.68	2	.025
1 or 0	26 (19.99)	9 (15.09)	20 (19.99)	4.22	2	.25
Total χ^2				11.90	4	.04
Pooled χ^2	36 (25.07)	10 (18.86)	23 (25.07)	9.09	2	.02
Interaction χ^2				2.81	2	.25

suggest that those who received more social support for their FSAs are more likely to be introjectors than those who received less.

Since the melody pairs were made as alike as possible, and neither tune was distinctly unpleasant, an unknown degree of uncertainty was created for the subjects, which may have affected the degree of introjection. In the cases in which the subjects found themselves in agreement with both the models on the two pairs of melodies, however, their preferences may have gained in social reality (Schacter, 1951). The greater certainty that probably resulted from the social reality may then have increased the tendency to introject.

Another test of this reasoning is provided by an examination of the differences in the number of DSAs made by introjectors with high and low degrees of social support for their preferences on the melodies. In order to test the last *ad hoc* hypothesis, it is first necessary to show that the two groups did not differ in the mean number of FSAs. A test of the differences in the mean number of FSAs yielded a t of only .15. The introjectors with high social support had a mean DSA score of 3.80, which was significantly higher than the mean of 2.96 for the other introjectors ($t = 1.56$, $p < .05$). The results support the *ad hoc* hypothesis that greater social support generates more introjection.

The second hypothesis asserted that the greater the number of FSAs, the more DSAs the individual projects into his concept of the model. In this case, the FSAs were computed by first calculating the total number of similarities in preferences with each model for both melodies and nonsense syllables and then subtracting the smaller from the larger figure. The degree of projection was measured by the number of times a subject assigned to a model the same preferences in names as her own. The first test of the second hypothesis was a tau correlation between the FSAs and the degree of projection on to the more similar model. This correlation was .22 ($p < .02$, one-tailed), supporting the hypothesis.

Another test of this hypothesis was made by dividing the subjects into three categories. The first, called projectors, were those whose preferences for names were granted more often to the model with whom they had been more similar in choices of melodies and nonsense syllables. Nonprojectors included subjects whose favored names were assigned to one model and to the other to an equal degree, whereas negative projectors were those whose preferences for names were more often assigned to models with whom they had least similarity on the music and nonsense syllables. These data are presented in Table 3.

The results show that the number of persons in the projector's category and in the other two categories were unlikely to occur by chance, again supporting the second hypothesis.

Considering FSAs based on the number of similar choices on melodies and on nonsense syllables separately, a tau correlation coefficient of the degree of similarity on musical preferences and the amount of projection on names was .22 ($p < .02$), but the correlation between the degree of similarity on nonsense syllables and the amount of projection on names was near zero and not significant. The degree of similarity in the choices on music, thus, accounted for almost all of the relationship between the first similar attribute and the amount of projection. The assumption that a DSA may function as FSA in subsequent projection is not supported.

Strength of Preferences

Well established FSAs may be more likely to lead to introjection than less well established ones. Persons who strongly preferred the tunes that were the basis of the FSAs might therefore be more likely to be intro-

TABLE 3

NUMBER OF PERSONS DEMONSTRATING DIFFERENT DEGREES OF PROJECTION

Projectors	Non-injectors	Negative Projectors	N	χ^2	df	p
42 (27.19)[a]	12 (11.62)	12 (27.19)	66[b]	16.57	2	.001

[a] These expected frequencies were calculated by determining the number of times out of 66 that one would expect by chance to get two series of 10 binary choices (predictions of subject for each model) which have an equal number of matches to a given series of 10 choices (own preferences). This gives the expected frequencies for the nonprojectors and the other two are calculated from the remainder.

[b] The total of subjects is 66 because 2 misunderstood the instructions and made no predictions. Two others had zero scores on the number of similarities on music plus nonsense syllables.

jectors than those who had weak preferences. The subjects rated the degree that they preferred the tunes they chose on a five-point scale in the final questionnaire. The number of subjects who had various degrees of introjection was determined for those subjects who stated that they had strong and weak preferences. Table 4 shows that a larger proportion of the persons who had strong preferences were in the introjection category, whereas a greater proportion of those with weak preferences were in the negative introjection category. The value of the interaction chi square indicates that the distribution of the persons in the groups with strong and weak preferences over the three introjector categories is significantly different. Thus, strongly based FSAs are more likely than weaker ones to lead to introjection.

Parallel results obtain for projection. In Table 5, it can be seen that those with strong preferences were projectors significantly more often than nonprojectors or negative projectors. Those with weak preferences were not projectors significantly more often, but the interaction χ^2 was not significant. In general, there is a tendency for those with strong FSAs to introject and/or project.

TABLE 4

EFFECT OF STRENGTH OF PREFERENCE FOR MELODY ON NUMBER OF PERSONS INTROJECTING

Strength of Musical Preferences	Introjector Categories			χ^2	*df*	*p*
	Introjectors	Nonintrojectors	Negative Introjectors			
Strong	29 (17.8)	9 (13.4)	11 (17.8)	11.09	2	.005
Weak	7 (6.9)	1 (5.2)	11 (6.9)	5.88	2	.02
Total χ^2				16.97	4	.02
Pooled χ^2	36 (24.82)	10 (18.36)	22 (24.82)[a]	9.17	2	.02
Interaction χ^2				7.80	2	.02

[a] One subject did not respond to the questionnaire items. This accounts for the reduction in frequency from 23 to 22 and the alteration of the theoretical frequencies from those previously used.

TABLE 5

EFFECT OF STRENGTH OF MUSICAL PREFERENCE ON NUMBER OF PERSONS SHOWING PROJECTION

Strength of Music Preferences	Projection Categories (Based Only on Musical Similarities)			χ^2	*df*	*p*
	Projectors	Nonprojectors	Negative Projectors			
Strong	31 (19.78)	10 (8.44)	7 (19.78)	14.91	2	.001
Weak	7 (7.42)	3 (3.16)	8 (7.42)	.08	2	.975
Total χ^2				14.99	4	.02
Pooled χ^2	38 (27.20)	13 (11.60)	15 (27.20)	9.93	2	.01
Interaction χ^2				5.06	2	.10

Perceived Similarity

To measure the subjects' awareness of their similarity to the models in the postquestionnaire, the subjects were given a triangle with the letters A, B, and C at the vertices (the subject was always C), and asked to place an "M" on the line connecting the most similar pair and an "L" on the line between the least similar. The first line of Table 6 shows that subjects had a strong tendency to perceive themselves as similar to the model with whom they agreed more on the melodies. When subjects with strong and weak preferences on music are compared, Table 6 shows that those with strong preferences tended to perceive themselves consciously as similar to the model with whom they agreed more on melodies. Those with weak preferences did not show a parallel trend significantly. (However, an interaction χ^2 was computed and not found to be significant.)

Do persons who are consistently similar to a given model on all types of stimuli perceive this similarity more than those whose similarity is not consistent? The subjects were divided into three groups. One, containing 36 subjects, was composed of those who had, on all three types of choice materials (melodies, nonsense syllables, and names) preferences that were similar to those of the same model. Of these 36 persons, 27 judged themselves as more similar to this model ($\chi^2 = 9.0$, $p < .05$). The second group contained those subjects who were more similar to the same model on two out of the three types of choice materials. There were 23 subjects in this group, of which 21 indicated that they saw themselves as most similar to this model ($\chi^2 = 15.7$, $p < .001$). The remaining 8 subjects (3 did not give sufficient data on the triad question for inclusion here) had similarities that were entirely inconsistent, in that they tended to match those of one model on one type of choice material, the other model on a second type of choice material, and both models

TABLE 6
NUMBER OF PERSONS CONSCIOUSLY AWARE OF SIMILARITY WITH MODELS

	Subjects Who Rate Selves as More Similar to P Who Has the More Matches with Subject's Musical Preferences.	Subjects Who Rate Selves as More Similar to P Who Has the Fewer Matches with Subject's Musical Preferences.	*N*	χ^2	*df*	*p*
All subjects	48	18	66[a]	13.70	1	.001
Subjects having strong musical preferences	37	11	48	14.08	1	.001
Subjects having weak musical preferences	11	7	18	.49	1	.50

[a] Omits four subjects; one subject had zero music similarity with models, and three subjects gave no judgments on triads.

equally on the third type. These subjects did not tend to see themselves as more similar to one model than to the other. It is clear that subjects who were similar to a model to a consistent or moderately consistent degree were likely to be aware of this similarity.

Finally, we ask if a subject who was aware of her similarity to the model on the choice materials generalizes the similarity to derived attributes. Subjects were asked to write a brief paragraph, on the basis of what they had learned during the experiment, describing each of the two models, and to write an additional one about themselves. From each paragraph those phrases were abstracted which contained traits, abilities, or characteristics not directly associated with music, sounds, or names. These were considered to be derived attributes that developed in the subjects' concepts of the models as the result of experience with them.

A panel of three judges determined whether the content of the paragraphs describing Model A or Model B was closer to the description each subject made of herself. Only those subjects who wrote three paragraphs and for which there was unanimous agreement among the three raters were used in this analysis. Of the 37 remaining, 26 rated themselves as more similar on the triad to that model whose description had been judged by the panel as most like their own. The other 11 persons saw themselves as most similar to the other model. These frequencies when compared to a chance expectation of equal distribution provide a χ^2 of 6.08 ($p < .02$). We conclude that persons who were aware of their similarity to others on the experimental materials most often attribute further similarity to others on separate matters.

DISCUSSION

We have learned from this experiment that the processes of introjection and projection may develop where there are FSAs between one's concept of himself and his concept of another person. Furthermore, there is some evidence to suggest that the greater the similarity, the more is introjection or projection likely to occur. The degree of social support for the FSAs and the strength of the preferences for those attributes increased the degree of introjection and projection of DSAs.

The results from the use of nonsense syllables with low associative value as attributes to be introjected were not always strongly significant. The failure of the variation in degree of similarity on nonsense syllables (DSAs) to contribute to the projection of names may have also been due to the low association value of the syllables. Because the syllables had little meaning for the subjects it may have been difficult for them to associate the syllables with other parts of their cognitive structure, and therefore not possible for them to associate preferences on nonsense syllables with other attributes of models. To be involved in the processes of introjection or projection as FSAs or DSAs attributes may need to be relevant or meaningful to the perceivers.

SUMMARY

Subjects working in isolation were led to believe that they were the third member of a three-person group and were asked privately

to write their preferences among eight pairs of brief tunes. After they had made each choice within each pair, they heard over earphones the two other choosers state their preferences. The subjects then stated their choices into a (dead) microphone. The two other choosers were in fact paid participants whose voices were transcribed. By appropriate means each subject came to perceive that her preferences were more similar to one of the two paid participants than the other. The subjects then made preferential choices from a series of pairs of nonsense syllables. In this instance, however, the subjects heard the selections of the other two persons before making each decision, but she was led to believe that her own selection could not be heard by the other subjects. It was found, as hypothesized, that the subjects tended to prefer the nonsense syllables chosen by the paid participant with whom they agreed more often on musical preferences, provided that the musical preferences were strong, or provided that the subjects received social support for their musical preferences.

Subjects were also asked to make choices from a series of pairs of girls' names and were asked to guess the preferences of the other two members. It was found, as hypothesized, that subjects assigned their own preferences on girls' names to the paid participant with whom they more often agreed on musical preferences.

The results were interpreted as giving support to a cognitive theory of identification.

REFERENCES

ASCH, S. E. *Social psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1952.

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

GLAZE, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

RAVEN, B. H. Sound influence on opinions and the communication of related content. *J. abnorm. soc. Psychol.*, 1959, **58**, 119–128.

SCHACTER, S. J. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

SHERIF, M. *The psychology of social norms*. New York: Harper, 1936.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 257-264

SIMILARITY TO A MODEL AND SELF-EVALUATION[1]

EUGENE BURNSTEIN
Michigan State University
EZRA STOTLAND
University of Washington
AND ALVIN ZANDER
University of Michigan

PREVIOUS research (Stotland, Zander, & Natsoulas, 1961) has shown that interpersonal identification can usefully be conceived in cognitive terms as a generalization of interpersonal similarities. The earlier study demonstrated that when a person finds that he has some characteristics, called first similar attributes (FSA), in common with another person or model, he tends to perceive himself as having other characteristics, called derived similar attributes (DSA), in common with the model. In the previous study, the derived similar attributes were trivial ones such as preferences for nonsense syllables and girls' names. One of the purposes of the present research is to determine whether generalization of similarity can occur where the derived attributes have more inherent importance to a person than nonsense syllables and girls' names. Primarily, it asks whether identification can influence a person's evaluation of himself. It also complements the exclusively cognitive emphasis of the previous study by inquiring whether an individual's desire to perceive himself favorably affects his identifying with a model.

The theory employed in the previous study divided the process of identification into two subprocesses, introjection and projection. Introjection was defined as the process by which the person changes his conception of himself to be more like his concept of the model, i.e., the DSA is located in his self-concept. On the other hand, projection is defined as the process by which he changes his conception of the model to be more like his concept of himself, i.e., the location of the DSA is in the person's conception of the model. In the present study, the main focus is on introjection. The first hypothesis replicates one in the previous study.

Hypothesis 1: The more similarities (FSAs) a person perceives between his conception of himself and his concept of another person or model, the greater is his tendency to introject DSAs into his concept of himself. This hypothesis refers to DSAs that are neutral with regard to the person's self-evaluation.

Given this view of introjection, it is plausible to assume that a person may employ the process instrumentally to acquire desirable DSAs. Such an assumption implies that a person who has few FSAs with a positive model (i.e., a model who has desirable attributes) is more likely to attempt to seek further evidence of similarities than one who has a similar number of FSAs with a negative model. Take a person who has only a few neutral FSAs between his self-concept and his concept of a model and who then discovers that the model has other desirable or undesirable qualities. The person's motivation to think well of himself, we assume, has a strong effect upon his cognition. If the attributes of the model are negative, the person tends to develop few, if any, additional neutral FSAs and may attempt to weaken those that already exist by changing attributes of himself, by distorting perceptions, or by attempting to deny that similarities exist. By doing so, he minimizes the probability of introjecting the negative qualities. On the other hand, if the model's attributes are positive, the individual should attempt to increase the number of FSAs by appropriately changing attributes of himself, distorting or denying, etc. It is, of course, assumed that persons have learned through past identification how introjection can be employed to acquire positive qualities in the self-concept.

If the person has a high number of neutral

[1] This research was financed by a Grant-In-Aid M-701 (C) from the National Institute of Mental Health, United States Public Health Service to the Research Center for Group Dynamics, University of Michigan, Alvin Zander, principal investigator.

FSAs with a negative model, he should be less likely to be able to reduce his similarity than one with few FSAs, since, we assume, it is more difficult to change many established FSAs. Likewise, if he has a high number of FSAs with a positive model, he may already have sufficient basis for introjecting desirable DSAs and therefore less need to increase his FSAs.

Hypothesis 2 then is: Given situations in which persons and models have few FSAs, a person who perceives a positive model should attempt to develop more FSAs with the model than a person who perceives a negative model.

We have, thus far, assumed that identification based on similarity can be used instrumentally to change self-evaluation. But the self-evaluation of a person may also be affected by the presence of FSAs that he has not himself established but occur because of factors beyond his control. The critical question here is whether a person's cognitive processes can lead him to acquire even undesirable attributes in his self-concept.

Hypothesis 3: The greater the number of FSAs between a person and a model, the more the person should raise his estimate of the degree he possesses a socially desirable attribute if the model is high in this attribute, and lower his estimate if the model is low in it.

METHOD

Young boys were addressed by a person in an interesting role—a deep-sea diver—taken by one of the authors. He told them of the abilities one must have in order to be a diver and of his own abilities, preferences, and experiences in his career. Two weeks prior to this talk the boys had completed a questionnaire in which they had been asked to rate their ability in a number of skills. The diver described his own ability in many of these same activities. For half of the boys he depicted himself as highly competent in these areas and for the other half he revealed that he was grossly incompetent. Within each of these two degrees of competence, the diver was presented as either highly similar or dissimilar to the boys. In a third treatment nothing was said concerning degree of similarity. After the diver's speech was finished, measurements were obtained from the boys concerning evaluation of their abilities in the same skills they had previously rated as well as other reactions necessary for testing the hypotheses. The experiment design is shown in Table 1.

TABLE 1
DESIGN OF EXPERIMENT
(*N*s in parentheses)

	Similarity Conditions		
	High Similarity Induced (High Sim)	Low Similarity Induced (Low Sim)	No Induction Concerning Similarity (No Ind)
Ability level of diver (model)	High (21) Low (19)	High (19) Low (20)	High (18) Low (20)

Subjects

The subjects were 118 sixth, seventh, and eighth grade boys in a small school located in a thinly populated suburb of a large industrial city.[2] Persons of these ages were chosen because they would probably have little information, contradictory or otherwise, about the abilities and preferences of deep-sea divers and because it was unlikely that their perceptions of themselves were so strongly fixed that they would be unchangeable.

Pre-experimental Measures

Baseline measurements were obtained 2 weeks before the experiment. All subjects chosen for the study were assembled in the school gymnasium. The administrators of the pretest (who did not take part in the subsequent experiment) were presented to the boys as representatives of the "district board for school planning." Because the school was in the process of building a new gymnasium, the boys were told that information was needed for planning purposes about what physical activities the boys could do well and what ones they did poorly; also, the board wanted some idea of what the boys liked to do and what kind of persons they were. To provide this information they were asked to complete a "Describe Yourself Questionnaire."

This questionnaire consisted of two sections. In one they rated themselves on eight-point scales in respect to nine skills: excellence in swimming, ability to hold breath under water, ability to judge distance, capacity for remaining calm in an exciting situation, capacity for doing complex tasks with hands and fingers, ability in throwing a baseball, ability in throwing a football, excellence in shooting baskets, and ability to high-jump. The first five abilities listed were those which would play a part in the experimental induction; the last four were inserted as "cover" questions.

The second section contained eight pairs of abstract sketches and six pairs of names of colors. The subjects were asked to state which member of each pair they preferred. Other simple and concrete pieces of information were requested in order that the subjects would believe that they were describing the kinds of persons they were.

Experimental Conditions

The diver met with small groups containing 10 to 15 boys. Among the 12 available groups, 4 were given in-

[2] The authors wish to express their deepest appreciation to Raymond Cypher, and Frank Rehamek, Atherton, Michigan, for their kind assistance in providing subjects for this experiment.

formation intended to generate an awareness that the subjects and the model were highly similar, 4 were given to understand that the boys and the model were not at all similar, and 4 were told nothing about the degree of similarity between the model and boys so as to allow spontaneous perception of different degrees of similarity.

In all groups the model was introduced to the boys as a member of the "University of Michigan Institute of Marine Salvage Engineering." The meaning of these terms was explained and they were told that the diver was going to speak about the abilities that are necessary in deep-sea diving. Before getting to the topic of his lecture, in the high and low similarity conditions, he made certain introductory remarks concerning either the degree of his similarity or lack of similarity to the listeners.

Induction of high and low similarity. In the high similarity (High Sim) condition the diver stated that he was interested in meeting with the group because at their age he was very much like them. He was born and raised in the neighborhood. He went to a school like this one. His father worked in the factory that employs most of the subjects' fathers. He participated in activities peculiar to boys in that area: swimming, boating, and fishing in a nearby lake. In fact, he went on, the principal was impressed at the similarity between the model's life and the lives of the subjects. This had led him to ask the diver to take a short test the boys had taken several weeks ago (the abstract drawings and colors). The principal was said to have been amazed at the similarity between the diver's choices and those of many, but not all, of the boys. He thought the boys would be interested in seeing this similarity. He then handed back to each boy the part of the premeasurement questionnaire that contained their choices among the drawings and colors and referred to it as a test of personality which accurately describes "the kind of person you are." On the same sheet, subjects were told, the principal had seen to it that the diver's answers were copied next the choices made by each subject. The alleged choices by the diver were inserted on each answer-sheet so that the diver appeared to have made the same choices as the subject in 12 out of the 14 pairs. After a brief period to allow study of this information, and for each subject to discover that he among all the boys was certainly very similar to the model, the papers were collected.

In the low similarity (Low Sim) condition the experimental sequence was exactly the same as in the High Sim except that an attempt was made to induce a perception of the diver as a person who was not similar to the boys. Thus, the diver stated that this group of boys was of interest to him because he at their age was very unlike them. He was born and raised in a densely populated, apartment-house neighborhood in San Francisco. He went to a very large junior high school in a very large building. His father was employed as a fisherman. The activities in which he engaged as a boy were peculiar to a large city in a congested area and unlike those of the boys in the audience. He then went on to explain how surprised the principal had been about his low degree of similarity with these boys and produced the previous "personality test" which, in this case, demonstrated that the model made choices different from the boys in 12 out of the 14 pairs.

No induction of high or low similarity. In this condition no attempt was made to induct a perception of similarity or low similarity between the boys and the model. The diver was introduced and he immediately began his talk without any introductory remarks concerning common or different upbringing, play activity, or personality.

After he had completed the part of his talk concerned with his abilities and before he went on to a discussion of the preferences he had developed as a diver, he paused to obtain information that was to measure the degree to which the subjects projected their own preferences into the model. The "personality test" was returned to each boy along with another unused copy of the questionnaire. The boys were asked to guess what choices the diver would make if he were taking the test. The diver then continued with the final portion of his speech.

Model's positive and negative attributes. At the outset of his lecture the diver produced a large bar-graph chart with the name of the "Maritime Institute" at the top and labeled "ability levels for divers of 1st, 2nd, and 3rd class." The speaker explained that divers fall into three classes and that their classification is determined by their ability in various skills. A first-class diver must be excellent in these activities, while a third-class diver need be only poor in them. The graph consisted of three large bars of grossly different heights that made the wide disparity between the ability of a top-level and low-level diver clearly visible. Eight skills were listed on the graph: dark vision, calmness, manual dexterity, plan of search, swimming, endurance of loss of air supply, endurance of small closed spaces, and distance judgment. Of these, it may be noted, five are the same as the skills on which the boys had previously rated themselves.

In the conditions in which the diver was to be a positive model he indicated that he was a member of the first class. In those cases where he was to be a negative model he stated that the third-class bar summarized his ability level. The body of his talk was a discussion of each of these skills. He pointed out why each of them was important in deep-sea diving, what the nature of an excellent and poor performance was, and by example and direct self-evaluative statements how well he was able to perform or not perform each of them. As a positive model he was consistently excellent in each skill and as a negative model he was consistently inept.

Statement of Personal Preferences

In order to test the degree to which neutral attributes of the model were introjected (for Hypothesis 1), the subjects had to be made aware of attributes of the diver that had no clear value for them. This was done by having the diver conclude his talk with a brief discussion about the unusual matters that different divers prefer: some divers prefer to descend from a ship, others from a dock; some prefer to dive in the morning, others in the afternoon; some would rather work in the Pacific Ocean, others in the Atlantic; and some would rather dive on a sandy floor, while others like a stony bottom. The diver stated his preference on each of these matters and admitted that he could not explain his preferences, that every diver simply has them regardless of his ability.

Post-experimental Measures

At the conclusion of the talk an experimenter who served as chairman told the boys that the Department of Maritime Engineering and Salvage was interested in how persons of their age would respond to questions similar to those asked of trainees in diving. He then gave them a questionnaire concerning three general areas: they estimated the degree to which they were similar to the model at present and the degree they would be like him in the future, they chose from among specified alternatives the working conditions they would prefer if they were divers, they rated their own proficiency in particular skills (the same ones they had rated several weeks earlier) as compared to other boys of their own age. Ratings were also made regarding the ability level of the experimental group. The boys were told that the diver would not see their responses.

The six experimental conditions were created, two groups in each, during a single school day, excluding communication between the groups during the period of data collection. All boys were assembled several days later and the experiment was explained and discussed.

RESULTS AND DISCUSSION

Check on Induction of Similarity

Rehearsals of the experiment showed that the degree of similarity the boys perceived between themselves and the model could not be measured meaningfully immediately after the inductions were made (prior to the main part of the diver's speech), since the boys simply reported back what they had been told. The alternative was to determine the degree to which subjects rated themselves as similar to the model after the diver's talk had been finished. By this time, however, they had also heard him describe his positive or negative qualities and his preferences, although these descriptions did not further emphasize high or low similarity. Thus, despite the events beyond the first inductions of similarity, subjects in the High Sim condition were expected to rate themselves as "like the diver" (now and in the future) more than those in Low Sim. Table 2 shows that subjects rated themselves as being more similar to the diver in the High Sim condition than in the Low Sim condition, whether the model was positive or negative in quality.

Introjection of Neutral DSAs

The first hypothesis is that persons who have a high number of FSAs with a model should introject more neutral DSAs than those who have a small number of FSAs with him. That is, persons in the High Sim condition should accept more of the diver's stated preferences for diving conditions as their own than do those in the Low Sim condition. Table 3 shows that the hypothesis was confirmed.

The second hypothesis was based on the assumption that persons want to think well of themselves. The prediction was made that subjects would foster the development of FSAs with a positive model where none or few were already present, and would hinder or weaken their development with a negative model in the same circumstances. In order to test this prediction, the experimental treatment was created in which there was no in-

TABLE 2

MEAN[a] DEGREE SUBJECTS RATE THEMSELVES AS SIMILAR TO MODEL IN HIGH SIM AND LOW SIM CONDITIONS

(*N*s in parentheses)

Condition of Similarity	Attributes of Model		
	Positive	Negative	Both Types of Attributes Combined
High Similarity	3.38 (21)	3.37 (19)	3.38 (40)
Low Similarity	2.03 (19)	1.50 (20)	1.76 (39)
Significance of difference between rows by one-tailed *t* test	$p < .01$	$p < .01$	$p < .001$

[a] Means are based on a combination of two ratings of similarity to the model, one concerning "now" and the other concerning the future.

TABLE 3

MEAN NUMBER OF MODEL'S PREFERENCES CHOSEN BY SUBJECTS IN HIGH SIM AND LOW SIM CONDITIONS

(*N*s in parentheses)

Condition of Similarity	Attributes of Model		
	Positive	Negative	Both Types of Attributes Combined
High Similarity	4.95 (21)	4.89 (19)	4.93 (40)
Low Similarity	4.11 (19)	4.15 (20)	4.13 (39)
Significance of difference between rows by one-tailed *t* test	$p < .01$	$p < .02$	$p < .01$

duction of high or low similarity (No Ind). Recall that the model within this treatment was negative for half of the subjects and positive for the other half. In this treatment, the preferences with respect to diving conditions are considered as potential FSAs. In the No Ind condition nothing was said concerning similarity of attributes between the subjects and the model, and subjects were free to treat the statements made by the diver concerning his preferences as attributes of their own so that similarity could be established.

The prediction then becomes: subjects with a positive model should develop many preferences similar to those of the diver while those with a negative model should develop few similar preferences. In addition, the subjects made choices on the abstract sketches and pairs of colors. They were asked to guess, in this treatment, what the diver's preferences might be. It then also follows from the second hypothesis that subjects with a positive model assign to him more of their own preferences than those with a negative model.

The data confirm both predictions. When the model was positive, the subjects accepted an average of 4.61 of his preferences as their own choices; when he was negative they accepted an average of 3.61 ($p < .02$ by t test). In assigning to the model their own preferences on the materials of the "personality test," 5.44 were granted to the positive model and 4.10 to the negative one ($p < .02$ by t test).

We conclude that where subjects have established few FSAs with a model and therefore are free to perceive or not to perceive FSAs as they wish, the motivation to think well of themselves determines the extent to which they develop FSAs. They establish more FSAs with models who appear to promise the likelihood of providing positive rather than negative attributes.

Self-Evaluation

The third hypothesis proposed that the positive or negative quality a person assigns to derived similar attributes in his concept of himself is a function of the positive or negative quality of the model. That is, subjects within the High Sim condition should raise their evaluations of personal skills (from the level stated 2 weeks earlier) where the diver has positive qualities, and lower them where the diver has negative qualities.

TABLE 4

MEAN SELF-EVALUATIONS OF ABILITIES BEFORE AND AFTER PERCEPTION OF MODEL IN HIGH SIM AND LOW SIM CONDITIONS

(*N*s in parentheses)

Condition of Similarity	Attributes of Model[a]	Mean Self-Evaluation			
		Before Perception of Model[a]	After Perception of Model[a]	Mean and Direction of Change[a]	Significance of Mean Change (One-Tailed t Test)
High Similarity	Positive (21)	3.71	3.90	.19 (a)	ns
	Negative (19)	3.77	3.49	−.28 (b)	.05
Low Similarity	Positive (19)	3.46	3.44	−.02 (c)	ns
	Negative (20)	4.02	3.41	−.61 (d)	.01

[a] There are no significant differences between means within these columns.

[b] The following cells in this column were compared by a one-tailed t test: a vs. b, $p < .02$; d vs. c, $p < .02$.

The data for the High Sim condition are in the top half of Table 4, and give some support to the hypothesis. Where the model was positive subjects increased their self-evaluations; where the model was negative, subjects lowered them. The difference between the changes in these two conditions is significant at the .02 level, and the change in evaluation in the negative condition is significant at the .05 level.

A corollary to the hypothesis would assert that the increase or decrease in the subjects' ratings of their personal abilities, depending upon the attributes of the diver, should be greater in High Sim than in Low Sim. A comparison of the results in the top half of Table 4 with those in the bottom half of that table provides a test for this hypothesis. The boys who had a positive model raised their evaluations of themselves to a greater degree in the High Sim treatment than in the Low Sim ($p < .02$), as predicted. Subjects in both High Sim and Low Sim who had a negative model lowered their evaluations of themselves significantly, but the drop was not significantly greater in the High Sim treatment than

in the Low Sim. This result is not in accord with the prediction. Furthermore, the decrease in self-evaluation, after hearing a negative model in the Low Sim treatment, was significantly great in itself, when it was expected to be small.

Why did the subjects in the Low Sim negative condition lower their ratings so substantially? One explanation is found in the fact that those in the Low Sim negative condition had unusually high self-ratings at the time of the pretest, and had farther to drop than those in other conditions when they rated these same abilities during the experimental session. Their mean amount of downward shift could, thus, be considerably larger by statistical artifact.

Subjects were therefore divided into three groups within each of the four experimental conditions according to their evaluations of themselves on the pretest (high, middle, and low). The mean amount of change in these groups for each condition may be seen in Table 5 which shows, first, that the subjects who rated themselves low or medium on the pretests did not manifest any significant changes in self-evaluation nor significant differences among changes. On the other hand, among subjects who rated themselves high on the pretest, there were differences in the amount of change. In the Low Sim conditions, there was significantly ($p < .05$) more drop of evaluation when the model was negative than when he was positive. In the High Sim condition, the difference in changes in self-evaluation under a positive and a negative model is of the same great order of magnitude as in the Low Sim condition, although not significant. In fact, when the High and Low Sim conditions were combined, the difference between the positive and negative model conditions is significant at the .001 level.

These significant differences indicate that in the negative model conditions, subjects who were high in pretest self-evaluation dropped markedly during the experimental session, either as a result of normal statistical regression or of the introjection of the negative traits of the model. Since the subjects who perceived a positive model appeared to have had a tendency to introject his attributes (showing little if any regression from their original higher self-evaluation) those perceiving a negative model may also have introjected the model's attributes. On the basis of the present data, it is not possible to choose between regression and introjection as explanations.

TABLE 5

MEAN CHANGES IN SELF-EVALUATIONS OF ABILITIES WITH DIFFERENT LEVELS OF SELF-EVALUATION ON PRETEST, IN EACH EXPERIMENTAL CONDITION (*N*s in parentheses)

Condition of Similarity	Attributes of Model	Level of Self-Evaluation on Pretest		
		High	Medium	Low
High	Positive	−.10 (6)	.03 (4)	.39 (11)
	Negative	−.62 (6)	.05 (7)	−.31 (6)
Low	Positive	.01[a] (7)	.10 (6)	−.16 (5)
	Negative	−.80[a] (9)	−.12 (6)	−.21 (5)

[a] Cells significantly different by a *t* test.

If the interpretation in terms of introjection is accepted, why did the subjects with a positive model in the Low Sim condition introject the positive traits of the model, when the theory postulates that such introjection can occur only on the basis of high similarity in attributes (FSAs)? A reasonable explanation is that the attempt to create low similarity for this group was not effective; they saw themselves as similar to the model, despite instructions which emphasized the lack of similarity. In fact, they rated themselves as significantly more similar to the model than did subjects who had a negative model, as appears in the bottom half of Table 6. Furthermore, 5 of these 7 subjects were above the median of the 20 subjects in the Low Sim positive in the degree to which they rated themselves as similar to the model. Finally, all 7 were above the median in the number of the model's preferences they accepted as their own.

Thus, these subjects may not have regressed to the mean because they established enough similarity with the positive model, and were stimulated by him to think at least well enough of themselves to overcome a tendency toward statistical regression in their self-ratings. Because they had rated themselves very high on the pretest, there was further

not much room for them to make an upward shift on the rating scale when evaluating themselves after perceiving themselves as similar to the positive model. The fact that they maintained their originally high self-ratings, while other groups with a negative model and initially high self-evaluations regressed, is consistent with Hypothesis 3, since their introjection of positive attributes was enough to hold their ground against tendencies to regress in self-ratings.

Why did the persons who had high evaluations of themselves on the pretest make themselves similar to a positive model in the Low Sim condition despite elaborate instructions concerning their lack of similarity? Let us assume that the high evaluations they originally ascribed to themselves are an indication of their high self-esteem. Prior research has shown that persons with high self-esteem have a strong tendency to protect their high opinion of themselves against any influences that might lower it, and that they have the confidence in themselves to do so (Stotland, Thorley, Thomas, Cohen, & Zander, 1957). We assume that this is also true here. Where a positive model indicated that he had little similarity with the subjects, only those with high self-evaluations on the pretest had the confidence to contradict this information from him and to see themselves as similar to him. This contradiction was provoked, we believe, by the efforts of the subjects with high self-esteem to maintain it through perceiving themselves as similar to the positive model. Subjects who were medium or low in self-esteem, on the other hand, did not appear to be able to contradict the experimental instructions.

TABLE 6

MEANS OF SUBJECTS' SELF-RATINGS OF SIMILARITY TO MODEL FOR SUBJECTS OF DIFFERENT LEVELS OF SELF-EVALUATION ON THE PRETEST, IN EACH EXPERIMENTAL CONDITION

(*N*s in parentheses)

Condition of Similarity	Attributes of Model	Level of Self-Evaluation on Pretest		
		High	Medium	Low
High	Positive	3.67 (6)	3.40 (5)	3.20 (10)
	Negative	3.50 (6)	2.00 (7)	3.33 (6)
Low	Positive	3.07 (7) (a)[a]	1.50 (6) (b)[a]	1.77 (6) (c)[a]
	Negative	1.28 (9) (d)[a]	1.33 (6)	1.77 (5)

[a] Cells significantly different by two-tailed *t* test; a vs. b ($p < .01$); a vs. c ($p < .01$); a vs. d ($p < .01$).

CONCLUSIONS

The results of this study provide additional support for the basic hypothesis of the theory of identification used here and already supported by the results of the study by Stotland, Zander, and Natsoulas (1961): that the presence of FSAs between a person and a model leads the person to introject into his self-concept some of the other characteristics of the model (DSAs), when these additional characteristics are neutral with regard to his self-evaluation. The results also suggest that, when the attributes of a model are desirable or undesirable, the person's level of self-esteem influences his tendency to introject these attributes. Persons with low self-esteem seem not to react to any marked degree, by not introjecting any attribute. Persons with high self-esteem, on the other hand, appear to be able to take advantage of available possibilities of establishing FSAs with models with desirable attributes, even when the establishment of such FSAs contradicts the implications of the information given by the experimenter.

The results also show that, when there is no obstacle such as contrary information from the experimenter, subjects establish FSAs with models who have desirable attributes more than with those who have undesirable attributes.

SUMMARY

An adult model was presented to groups of grade school children as being either highly similar or of little similarity to the children in background and other attributes. To half of the groups he described himself as being high in abilities necessary for deep-sea diving. To the other half he described himself as being deficient in these abilities. He also described a number of his preferences relevant to deep-sea diving. The subjects evaluated themselves on these same abilities both before and after the experiment, as well as indicating their own preferences relevant to deep-sea diving.

It was found that the subjects who were told that the diver was highly similar to them in background accepted his preferences relevant to deep-sea diving, more so than subjects who were told that he was not similar to them. It was also found that subjects who appeared to be high in self-esteem made themselves more similar in preferences to a positive model when he had told them that he was not similar to them in background attributes. Apparently as a result, these subjects were able to maintain their evaluation of their own abilities which they had made in a pre-experimental test. Subjects with high pretest self-evaluations, when presented with a negative model, regressed toward the group's mean self-evaluation from their high pre-experimental evaluation.

In another condition of the study, no attempt was made to induce a perception of high or low similarity of the subjects to the model. The subjects accepted as their own the preferences of the able model more than the preferences of the less able one, and projected more of their own preferences onto the model when the latter was able than when he was less able.

REFERENCES

STOTLAND, E., THORLEY, S., THOMAS, E., COHEN, A. R., & ZANDER, A. The effects of group expectations and self-esteem upon self-evaluation. *J. abnorm. soc. Psychol.*, 1957, **54**, 55–63.

STOTLAND, E., ZANDER, A., & NATSOULAS, T. Generalization of interpersonal similarity. *J. abnorm. soc. Psychol.*, 1961, **62**, 250–256.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 265-274

IDENTIFICATION AND CHANGES IN PREJUDICE AND IN AUTHORITARIANISM[1]

EZRA STOTLAND AND MARTIN PATCHEN[2]
University of Washington *University of Michigan*

PREVIOUS studies have shown that it is possible to reduce a person's degree of prejudice by communicating to him—in part via case histories—about some of the emotional, irrational bases of prejudice (Katz, Sarnoff, & McClintock, 1956; Stotland, Katz, & Patchen, 1959). Since the acquisition of insight involves a change in the person's conception of himself, any process that leads to a change in self-conception might be employed to give insight into the emotional bases of prejudice and, thus, reduce it. One process of changing self-conception is identification with another person or model. The results obtained in these studies may, thus, have followed from the fact that the person in the case history served as a model with whom the subjects identified. The study reported here attempts to establish that such identification can facilitate the acquisition of insight and the consequent reduction in prejudice.

The proposition that identification leads to changes in self-conception finds support in the work of Stotland, Zander, and Natsoulas (1961). When a person has an attribute in his concept of himself that is similar to one in his concept of a model, these authors term this attribute a first similar attribute (FSA). It is assumed that persons tend to conceive of themselves as possessing the same (or a similar) attributes as those that they perceive to be associated with the FSA in their concept of the model except when veridical self-perception would be contradicted. Such induced attributes are termed derived similar attributes (DSAs). The type of identification that leads to the acquisition of these traits is termed introjection. The strength of a person's tendency to introject is assumed to be partly a function of the number of FSAs in his concepts of himself and of the model.

This process may be one of those by which a person accepts insights about his ego defenses and about how they determine his prejudicial attitudes. Suppose that a person finds that he has certain common qualities (FSAs) with another person. Suppose also that the person conceives that this model is prejudiced against certain minority groups because of the functioning of such ego defenses as projection, compensation, displacement of hostility, etc. He would be expected then to develop in his self-concept an awareness that he himself also has the attributes of engaging in these ego defenses. That is, he gains insight.

Self-insight about the defensive component of prejudice, once acquired, may be expected to lead one to reduce his defensiveness and his prejudice, on the assumption that people in our culture are trained to think of irrationality and devices like scapegoating as "bad." Furthermore, the materials that our subjects read concerning the relationship between emotions and attitudes clearly implied that defensive reactions and hostility to other groups do not reflect a person's best qualities.

Thus, the first hypothesis is: If a person conceives of another person as being prejudiced toward minority groups for ego defensive reasons, and if he also conceives of himself as having several objective characteristics (FSAs) that are similar to those of his concept of the other person (the model), he should tend to reduce his prejudice toward minority groups more than if he conceives of himself as having fewer objective characteristics similar to those of the other person.

The general hypothesis was not, however, expected to hold equally for all persons. Individuals who do not actually engage in the same defenses as the model would be expected

[1] This research report draws its data from the program of research on attitude change under the direction of Daniel Katz. This research was supported in part by the United States Air Force under Contract No. AF 41 (657)-82, monitored by Crew Research Laboratory, Air Force Personnel and Training Research Center (deactivated in March 1958).

[2] The authors wish to express their indebtedness to Daniel Katz for his help, but the responsibility for the analysis and the interpretations reported in this paper are theirs. We are also indebted to John J. Jochem for his help in the administration and analysis of the data.

to reject the potential DSA of using these defenses, as inconsistent with their own actual behavior. In the light of the work of Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) persons low in authoritarianism should react in this fashion.

Since the DSA of engaging in defensive behavior would apply to the high authoritarians, such persons might be expected to conceive of themselves as also having this attribute. However, the additional factor of strong motivation to accept or reject the DSA must also be considered. The very strength of high authoritarians' defenses bespeaks a very great need for protection against anxiety. They would probably tend, then, to react defensively to reject any information that might lead to an insight. Both those who are very high or very low in authoritarianism would therefore be expected to resist the conception of themselves as behaving defensively. However, persons who are moderate in authoritarianism should be less subject to the forces on persons at either extreme.

Accordingly, the second hypothesis is that when persons are given information about a model whose prejudice is based on ego defensive reactions, those who are moderate in authoritarianism should tend more than those low or high in authoritarianism to reduce their prejudices toward minority groups.

As a (not necessarily conscious) cognitive process, the process of gaining insight into defense mechanisms and their relationship to prejudice resembles problem solving, since both entail a reorganization or a change of meaning of given attributes of entities. In problem solving, a typical finding is that an extended period of time, called incubation, must elapse between the presentation of the problem and the attainment of the solution (Woodworth & Schlosberg, 1954, pp. 838–891). During incubation, the person is not usually conscious of any cognitive work he is doing on the problem. If the process of gaining insight into oneself and problem solving are in fact similar, a period of time should be required before the significance of the insights is fully brought to bear. Thus, the effects of reading the case history should become evident only some time after the case history has been read.

Thus, the third hypothesis predicts a "sleeper effect"; i.e., that the losses in prejudice hypothesized in the first two hypotheses should not occur until after the passage of a considerable period of time.

METHOD

Subjects and Setting

The subjects for this study were 85 female college students at a midwestern university. The experiment was performed as part of regular class work in three meetings of undergraduate courses in introductory psychology and introductory social science. The experimental sessions were administered by the staff of this study rather than by the classroom instructors.[2]

Procedure

The over-all design of the experiments was as follows. *First session:* Measure of personality and attitudes toward minority groups obtained from all subjects. *Second session:* Three or 4 weeks later, an attempt was made to give the subjects insight into the emotional basis of their attitudes. This was done under two conditions: Where the person in a case history illustrating the relation of emotions and attitudes is someone very much like the subject in objective characteristics (FSAs with model very high), and where the person in the case history is not very similar to the subject in objective characteristics (FSAs with model very low). *Third session:* Immediately following the experimental induction and again 4 weeks later, attitudes toward minority groups were remeasured. In addition, the personality measures were repeated at the third session. No formal control group was included in the design since previous experiments in this program (Katz, Sarnoff, & McClintock, 1956) have demonstrated that attitudes *can* be changed through the self-insight method. The present study is concerned mainly with the problem of *why* such changes take place. These three sessions will now be described in more detail.

First session: Each subject was given a booklet that contained scales of personality and of stereotypes about Negroes. The Negro stereotype measure was the 21-item Likert scale used by Katz et al. (1956). Both favorable and unfavorable items were included. A person's score consists of the sum of the ratings of agreement with the favorable items. Thus, the lower the score, the less prejudice. The personality measures used were the 30-item F Scale used by Adorno et al. (1950); and a Rationality scale (to be described), which was intended to measure the tendency to view oneself and the world in a rational and objective way. In the first session, all subjects received identical question booklets and were given all the time they needed. They were informed that their answers would be completely anonymous.

[2] The authors are indebted to the following instructors for their generous cooperation in this study: Thomas Cohen, Robert Daugherty, Leonard Moss, Robert Fulton, Evangeline Norton, and Robert Rhodes.

Second session: The second session took place approximately 4 weeks after the first. At this session, each subject received one of two different types of booklets, realizing the two experimental conditions. The booklets were passed out in the classrooms from a pile in which both types were alternated. Thus, both experimental conditions were administered simultaneously in the same classroom.

General interpretive section. The two types of booklets had identical first sections. In all cases, subjects read first a five-page, double-spaced selection entitled, "Emotions and Attitudes." This section described in simple, layman's terms the ways in which emotional difficulties may lead to antiminority attitudes through such mechanisms as scapegoating, projection, and compensation. This selection was the same as the first part of the "interpretation" materials used by Katz et al. (1956).

Case history. The general statement about how attitudes are related to emotions was followed by a "case history" of a girl named "Carol." Subjects in the high similarity condition (High Sim) read about a person similar to themselves in objective characteristics, while subjects in the low similarity condition (Low Sim) read about a dissimilar person. Since we wish to see the effects on attitude change of different degrees of identification, separate from the effects of other variables like mobility aspiration, the two fictional "Carols" were described in terms which made them about equal in social status.

Background data on all subjects were obtained at the first administration so that facts about "Carol" could be tailored to be either like or unlike the subjects. Since subjects in some classes had slightly different backgrounds from those in other classes, it was necessary to have different descriptions of a similar Carol and different descriptions of a dissimilar Carol to fit the modal background of persons in each class.

Similarity condition. In the similarity condition (Sim), Carol was described as an American girl attending the University of Toledo. Her major subject in college, her vocational goal, her part-time job, her father's occupation, her recreations (e.g., going to the movies), etc. were all described in each class as closely similar to those of the subjects in that class. Subjects in each class fortunately tended to be homogenous on these characteristics.

Low Similarity condition. In the other condition, "Carol" was described as a student at "Canterbury University, located in the agricultural area of New Zealand." Her father's occupation, her part-time job, her major subject, her vocational goal, and her recreations (e.g., watching cricket matches) were all made alien to the average student at this university.

Following the description of "Carol" the rest of the case history was almost identical for both experimental conditions. For both "American Carol" and "New Zealand Carol," the relation of personal frustrations to antiminority attitudes was traced. Examples given of the prejudice of "American Carol" were those against Negroes, Mexicans, and Turks, while "New Zealand Carol" was described as disliking Negroes, Indonesians, and Turks.

Negro Attitude scale. Next all subjects filled out a scale measuring negative stereotypes of Negroes and feelings about behavior toward Negroes. The Stereotype scale consisted of 17 items matched to items of the Stereotype scale used in the first session, and was found to correlate .84 with the original scale when tested on another sample of college students. The purpose of changing the scale was to meet the problem of monotony in answering the same scale at all three sessions and the danger that subjects might develop stable responses to the items after two administrations. In addition to the matched items, 4 of the original 21 items were repeated to provide a constant attitude measure.

Third session: In the third session, approximately 4 weeks after the second, the same personality and attitude scales that had been administered in the first session were repeated.

Analysis of F Scale

In the analysis of the data, only a subscale of the F Scale, termed the Defensiveness subscale (McClintock, 1958), was used, since it purported to measure the tendency to employ the defenses with which this study is concerned. McClintock presents evidence suggesting that the personality syndrome of ego defensiveness operates in a different fashion in attitude change from the syndrome defined by the rest of the items in the F Scale.

Rationality Scale

The Rationality scale administered during the first session consisted of the following items answered by the Likert method of indicating degrees of agreement or disagreement.

1. What happens to us seems to be determined a lot by a fate over which we have little control.
2. There have been many times when I've felt jealous of individuals who have personal qualities which are better than my own.
3. If a person is in a position to help me in some way, I often act more friendly toward him than I really feel.
4. When a new idea seems to be logical but most people feel that it somehow doesn't make sense, you can be pretty sure there's some hidden flaw in it.
5. If I have a heated argument with someone, once it's over I forget all about it and don't hold any grudge against the other person.
6. When a custom isn't as efficient as possible, it should be changed—even though it is traditional and people are used to doing things that way.
7. I usually dislike someone if he disputes my opinions.
8. It is sometimes possible for a person to know what some close relation is thinking or that something awful is happening to him, even though he is far away.

This scale was designed to measure a person's tendency to judge the world and himself in a hard-headed, factual manner. Four of the items (1, 4, 6, and 8) concern the extent to which he judges objective

phenomena in a logical, nonemotional way. The other four items (2, 3, 5, and 7) concern the extent to which he views himself in an objective rational way. These latter items are statements true of almost all people; disagreement with 2, 3, and 7 and agreement with 5 entail the denial of fact and a tendency toward self-deception.

RESULTS

Differences between Conditions

The first hypothesis asserts that among persons who perceive another individual ("model") to be prejudiced toward minority groups for ego defensive reasons, those who have several objective characteristics that are similar to the model's should become less prejudiced toward minority groups. Before this hypothesis can be tested, High and Low Sim subjects must be shown not to differ significantly in their mean levels of prejudice at the first, pre-experimental administration. This requirement was met, the mean Negro stereotype score being 7.42 for High Sim and 8.41 for Low Sim.

The hypothesis leads to the prediction of greater reduction in prejudice in the High Sim condition than in the Low Sim condition. The prediction is not supported, since the mean changes of stereotype from the first to the third sessions for the Sim and Low Sim conditions, -1.70 and -2.30, respectively, do not differ significantly. (Negative numbers mean a reduction in prejudice.) In addition neither of the groups showed a significant mean change in prejudice from the first to the third sessions.

A more interesting picture emerges, however, when the mean changes in negative stereotype are considered in relation to the original degree of prejudice that the subjects manifested at the first session. Table 1 shows for each condition the mean changes in total stereotype scores separately for subjects high and low in prejudice at the first administration. Subjects low in original degree of prejudice show some tendency to become *more* prejudiced in the High Sim condition. These subjects show a mean change in a direction opposite to that for subjects in the Low Sim condition who started at the same level of prejudice. (This difference is significant at the .10 level.) Moreover, subjects in the High Sim condition are the only group who have a

TABLE 1

MEAN CHANGE[a] IN TOTAL NEGRO STEREOTYPE SCORE FROM FIRST TO THIRD SESSION IN RELATION TO ORIGINAL DEGREE OF PREJUDICE
(Number of Cases in Parentheses)

Original Degree of Prejudice	Experimental Condition	
	Low Sim	High Sim
Low Prejudice	-1.55 (21) A	3.65^{**} (30) B
High Prejudice	-3.70 (14) C	-1.25 (20) D

Note.—Significance of differences between cell means by two-tailed t tests: A vs. B, $p < .10$.

[a] Positive score means change to more negative stereotype.

** Indicates change from position at first session significant at .07 level, two-tailed t test.

mean change in stereotype from their initial position that approaches significance ($p < .07$). In short, the subjects who were low in prejudice show a tendency to become more prejudiced when there is similarity between themselves and the prejudiced person in the case history. In contrast, no differences between conditions or significant changes in either condition were found among subjects who were high in prejudice at the first session. The results of Table 1 are contradictory to Hypothesis 1. This shift among the relatively unprejudiced subjects cannot be explained solely as a regression effect since it occurred only in the High Sim condition.[4]

Authoritarianism and Change in Prejudice

Confirming the original studies of authoritarianism by Adorno et al. (1950) for the F Scale although only the Defensiveness subscale of the F Scale was employed, the correlation with anti-Negro prejudice is positive ($r = .27$, $p < .02$).

We expected that persons who were medium in defensiveness would change further away

[4] It is also possible to interpret these differences as showing that the low prejudiced in the Low Sim condition were somehow prevented from showing high prejudice scores as a result of normal regression. The failure of this group to become more prejudiced may have resulted from pressure from the experimenters and the regular course work toward less prejudice. On the other hand, the Sim condition appears to have contributed to these pressures.

from prejudice than would those who are neither high nor low in defensiveness. An analysis of the relation between level of authoritarianism and change in Negro stereotype, separately for those who started with little prejudice and for those who started with strong prejudice, show that this hypothesis is not supported. None of the changes was significant for any of the three groups on the defensiveness dimension, regardless of their original degree of prejudice. Nor was there any significant difference among the groups in the amount of change.

The obvious next question is whether the subjects' position on the Defensiveness scale had any influence on their reactions to the differences in the experimental conditions. Table 2 shows that people who are low in defensiveness and who experience the High Sim condition become significantly more prejudiced by the third session than they were at the start of the study. This group changes significantly more than those from the Low Sim condition who are also low in defensiveness, and more than subjects with medium or low defensiveness in the Sim condition. It is those who score *low* in defensiveness and who are in the Sim condition who become more prejudiced.

Since the low-defensive persons become more prejudiced in the High Sim condition, it might be expected that they would also acquire other authoritarian characteristics possessed by the model. This expectation is confirmed by the data of Table 3, which show that low-defensives in the High Sim condition scored significantly more defensive at the third session than they had at the first session. This change toward greater defensiveness differed significantly from changes among those who were also low defensives but were in the Low Sim condition.

TABLE 2

MEAN CHANGE[a] IN TOTAL STEREOTYPE SCORE FROM FIRST TO THIRD SESSIONS IN RELATION TO DEFENSIVENESS SCORE

(Number of Cases in Parentheses)

Position on Defensiveness Subscale	Experimental Condition	
	Low Sim	High Sim
Low	−1.35 (18) A	7.70*** (14) B
Medium	−4.45 (7) D	−.80 (25) E
High	−2.50 (10) G	−.25 (11) H

Note.—Significance of differences between cell means by a two-tailed t test: A vs. B, $p < .05$; B vs. E, $p < .05$; B vs. H, $p < .01$.

[a] More positive score means change to more negative stereotype.

*** Indicates change significant at .05 level in two-tailed test.

TABLE 3

MEAN CHANGE[a] IN DEFENSIVENESS SUBSCALE OF F SCALE[a] FROM FIRST TO THIRD IN RELATION TO INITIAL DEFENSIVENESS SCORE

(Number of Cases in Parentheses)

Original Position Defensiveness Subscale	Experimental Condition	
	Low Sim	High Sim
Low	1.95 (18) A	8.90**** (14) B
Medium	7.15 (7) C	−1.80 (25) D
High	−3.50 (10)	−4.55 (11)

Note.—Significance of differences between cell means by a two-tailed t test: A vs. B, $p < .10$; C vs. D, $p < .10$.

[a] Positive score means change to greater defensiveness.

**** Indicates change significant at .001 level in two-tailed test.

Rationality Scale and Attitude Changes

Since the Rationality scale was used in the study primarily for exploratory purposes, the first step is to discover the relationship between original degree of prejudice and degree of rationality. The mean score on the Negro Stereotype scale at the first administration was 4.28 for those high in rationality and 4.51 for those low, when subjects are divided at approximately the median (difference not significant). Level of rationality can therefore be regarded as independent of the initial degree of prejudice.

In Table 4 are presented the mean changes in Negro stereotype scores from the first to the third conditions for subjects high and low on the Rationality scale. Subjects who are low in rationality tended to shift ($p < .10$)

TABLE 4

MEAN CHANGE[a] IN TOTAL STEREOTYPE SCORE FROM FIRST TO THIRD SESSIONS IN RELATION TO RATIONALITY SCORE

(Number of Cases in Parentheses)

Position on Rationality Scale	Experimental Condition	
	Low Sim	High Sim
Low Rationality	−3.50* (20) A	2.95 (31) B
High Rationality	−0.65 (15)	−0.35 (19)

Note.—Significance of difference between cell means by two tailed *t* test: A vs. B, $p < .05$.

[a] More positive score means change to more negative stereotype.

* Indicates change significant at .10 level in two-tailed test.

TABLE 5

MEAN CHANGE IN TOTAL STEREOTYPE SCORE[a] FROM FIRST TO THIRD SESSIONS IN RELATION TO RATIONALITY AND INITIAL PREJUDICE

(Number of Cases in Parentheses)

Rationality Score	Initial Low Prejudice		Initial High Prejudice	
	Dissimilarity Condition	Similarity Condition	Dissimilarity Condition	Similarity Condition
Low Rationality	−5.10*** (12) A	5.70*** (19) B	−1.10 (8) C	−1.35 (12) D
High Rationality	3.65 (9) F	0.20 (11) G	−7.15 (6) H	−1.10 (8) I

Note.—Significance of difference between cell means by two-tailed *t* test: A vs. B, $p < .05$; F vs. H, $p < .10$; B vs. D, $p < .10$; A vs. F, $p < .05$.

[a] Positive score means more negative steotype.

*** Indicates change in cell significant at .05 level. Two-tailed test used where change is positive.

toward less prejudice in the Low Sim condition, while those in the High Sim condition showed a slight tendency to shift in the other direction, although neither shift is statistically significant. The difference between these groups is significant at the .05 level. On the other hand, subjects who are high in rationality do not appear to have been affected by the differences between the High Sim and Low Sim conditions. These results might suggest that it was the Low Sim condition that led to the greatest change in Negro stereotypes and not the High Sim condition, as in the previously reported results. However, the data in Table 5 help to clarify this issue.

Table 5 shows the relationship between attitude change and rationality separately for subjects high and low in original degree of prejudice. We see that, for subjects who are high in initial prejudice, neither their degree of rationality nor the experimental condition they experienced makes a significant difference in how much they changed in Negro stereotype. On the other hand, the subjects who are low in original degree of prejudice and who are also irrational are the ones most affected by the difference between the conditions. In the High Sim condition, they change significantly toward more prejudice, while in the Low Sim condition they show a significant change toward less prejudice. Thus, for those low in rationality and low in original prejudice, both conditions were leading to significant changes—in opposite directions. The difference between the conditions in the mean degree of change is significant at the .05 level.

An additional finding shown by Table 5 is that in the Low Sim condition, among subjects low in original prejudice, irrational persons changed significantly more in the direction of lower prejudice than did rational ones. Although this finding is not inconsistent with the previous finding, it is difficult to understand why these irrational subjects should tend to be influenced toward less prejudice in the Low Sim condition.

The correlation between the Rationality scale and the Defensiveness subscale of the F was found to be −.26 ($p < .02$), showing some tendency for the low rationals to be high in defensiveness. Since it was found that low defensives and low rationals were more susceptible to identification, these dimensions of personality appear to have different significance, although they are related.

Sleeper Effect

In order to test the hypothesis predicting a sleeper effect four items concerning Negro stereotypes were repeated at all three sessions. A sleeper effect would be indicated by the absence at the second session of changes that appeared at the third. All of the significant changes appearing at the third session were related to initial position on prejudice or on a

personality dimension. Thus, the sleeper effect can be checked by examining changes from the first to the second sessions for persons at different positions on these variables.

We have seen that in the High Sim condition, subjects who were low in initial prejudice, low in defensiveness, or low in rationality all changed significantly (toward greater prejudice) from the first to the third sessions. None of these groups shows any significant change from the first to the second sessions, in keeping with a sleeper effect; nor did the changes they showed from the first to second sessions differ from those of any other group. These findings hold for both the total stereotype scales and the items duplicated at all three administrations.

However, subjects high in defensiveness in the Low Sim condition showed a significant drop in prejudice at the end of the second session when prejudice was measured by either the total scale or the duplicated items. This drop was evidently dissipated by the third session, since none of these subjects changed significantly from the first to third sessions. Although this finding is contrary to the sleeper effect hypothesis, it is not inconsistent with the changes that were effected by the experimental manipulations over the entire time period of the study—i.e., a reduction or no change of prejudice in the Low Sim condition as compared to an increase of prejudice in the High Sim condition.

DISCUSSION

The first hypothesis of the study was that similarity between a subject and a prejudiced model, coupled with an explanation of the defensive basis of the model's prejudice, would lead the subject to gain insights into the defensive basis of her own prejudice, and, thus, to reduction of prejudice. The results, however, did not support this hypothesis. They indicated instead that persons of little prejudice tend to become more prejudiced when they find they have objective characteristics in common with a bigoted model. On the other hand, subjects who were already prejudiced did not change. While these results are unexpected, they can still be described in terms of our theory of identification.

We assumed that when a person conceives of himself and a model as having objectively similar characteristics, the person should tend to conceive that he is like the model in other respects called derived similar attributes (DSAs). In the present experiment, we attempted to induce the subject to see herself as similar to the model in the DSA of prejudice resulting from the use of defense mechanisms. Having perceived herself as like the model in this "undesirable" DSA, the subject was expected to want to change her own behavior to eliminate the undesirable DSA. The unexpected finding that certain subjects became *more* like the model in prejudice and in other related characteristics indicates that these subjects perceived themselves as like the model in the DSA of prejudice but did not take the next step of trying to change this trait. The question arises as to why these subjects did not try to eliminate the attributes that were clearly implied by the entire experimental situation to be "bad."

There are at least two ways of explaining this finding, one more directly motivational in orientation, the other more cognitive. From the standpoint of social supports for prejudice, in the first place, the High Sim condition may have inadvertently acted to make prejudiced attitudes legitimate. At least some of those subjects who initially scored low on prejudice may have had a good deal of latent anti-Negro feeling—held in check, perhaps, partially by the notion that other college level people in an intellectual climate of "tolerance" did not share such prejudiced views. The presentation of a vivid story about a person like themselves who had such anti-Negro prejudices may have served to give these subjects reassurance that they were not, after all, alone in these "wrong" feelings. Labeling these feelings as unrealistic may have been less important than the idea that they were shared by persons like themselves. The experimenters' fairly clear, though implied, opposition to prejudice may have been rendered relatively ineffective by the facts that they were persons of a very different status and strangers (from another school) to boot.

In this light, it becomes easier to understand the rather surprising finding that the subjects low in defensiveness were most susceptible to becoming more prejudiced, since

Berkowitz and Lundy (1957) have shown that persons low on the F Scale tend to be susceptible to conformity pressures from peer groups, i.e., groups with characteristics similar to the subject's. Furthermore, as indicated in the original report by Adorno et al. (1950, p. 771ff.), some persons are low on the F Scale only for reasons of conformity to certain liberal group norms and not for personality reasons. Thus, although some low defensives may not be disposed to prejudice by personality, their susceptibility to peer group influence makes them more susceptible to accepting the prejudiced norms of some peer groups. The particular susceptibility of the irrationals is consistent with this general explanation. To accept the prejudiced idea of peers even though their unrealistic basis is clearly explained is, after all, quite irrational behavior.

The cognitive explanation of the findings retains some of the original thinking about identification. Since subjects were made to perceive first similar attributes with the model, they have strong tendencies to perceive themselves in a manner consistent with the FSAs. In the present case, consistency is achievable mainly by perceiving themselves to possess the derived similar attributes (DSAs) that were associated with the FSAs. The subjects probably could not easily employ any other way of achieving consistency within the context of the present study. Achieving consistency by adopting greater prejudice was unopposed in the present study by any strongly aroused motivation to behave rationally and reasonably. In a previous study (Stotland, Katz, & Patchen, 1959), the experimental arousal of such motivation to be rational and reasonable was found to be essential to the lowering of defensively based prejudice by means of communicating information that could provide the basis for insight into the defenses. Since no similar effort to arouse the subject's motivation to be rational was employed in the present study, the subjects could become like the model even in ways that were negatively evaluated by the experimenters as irrational and unreasonable. The fact that some of these changes were observable only after a period of time argues for this cognitive explanation, since cognitive changes probably take time to occur.

This cognitive explanation is also consistent with the finding that the subjects low in rationality were most prone to identify. Since such subjects were assumed to tend not to perceive themselves objectively, they should therefore be more likely than others to shift their self-perceptions to be like those of a model. To accept the prejudice of the model, they had to disregard some of their own previous feelings and to magnify the prejudiced feelings shared with the model. It may be noted that the original, unsupported prediction of a shift toward less prejudice in the Sim condition was based on the assumption that the identifiers would have some tendency to reject irrational forms of behavior. The hypothesis may have failed to be confirmed in part because it was the unrealistic, irrational persons who tended to change their self-perceptions readily to be more similar to a model. Therefore, the identifiers were persons who would be expected to place minimal pressures on themselves to become more reasonable.

The cognitive explanation can also be reconciled with the finding that the low defensives became more defensive and more prejudiced in the High Sim condition. Adorno et al. (1950) suggest that persons low in F relate easily, warmly, and tolerantly with others. The low defenders may accordingly have a strong tendency to respond to, and identify with, other persons and to be more susceptible to influence by them. Such a tendency toward identification may extend even to those models who have harsh, aggressive attitudes—especially if the low defenders' tolerance is superficial. Highly defensive persons, on the other hand, may tend to project their faults onto others and then reject them. It is as if the high defenders relate to others in terms of their own needs, and, thus, are less likely to change through identification with others.

Regardless of which explanation, cognitive or motivational, should ultimately prove more valid, a question remains as to the exact nature of the changes on the part of subjects who identified with the model. In view of the relatively superficial experimental inductions, the shift in scores on the F Scale or the prejudice scale cannot reasonably be thought to represent a profound change in personality

functioning. On the other hand, the fact that changes in F Scale scores occurred when only antiminority attitudes were the direct target of change indicates that some generalized effect took place. Most likely, a perception of oneself as being tough or severe in outlook, like the model, was acquired. Such a characteristic could be reflected by the subjects not only in their expressed degree of prejudice, but also in their acceptance of the F Scale items, a number of which involve harsh toughness in evaluation of people. Thus, the acquired characteristic can be reasonably understood as a change in self-perception, not necessarily a change in overt behavior. If it was toughness that was accepted, the attempt to develop genuine insight among the subjects into the irrational basis of prejudice clearly failed. Since an entirely different process occurred the study does not provide any data on the question of whether insight actually does lead to a reduction of prejudice. Perhaps if the subjects had been induced to identify with the model at the same time that their motivation to be rational had been aroused, the study would have had different results.

The second hypothesis of the study was that persons medium in authoritarianism would be most susceptible to being changed by experimental inductions. This hypothesis was not supported but the conditions necessary for its confirmation were not achieved, since the experimental inductions did not result in the acquisition of insight. The findings concerning defensiveness were that, in the High Sim condition, the low defensives changed their expressions of general attitudes in the way that would make them more like the model, i.e., more authoritarian and more prejudiced.

The third hypothesis of a sleeper effect was in general confirmed. None of the long-term changes observed were manifest at the end of the second session. Since cognitive reorganization presumably takes time, these results are in keeping with the interpretation that the acquisition of more prejudice and defensiveness was due to a change in the self-concept.

The results of this study have some important social implications. Many communications—e.g., in movies and television—vividly portray the activities and sentiments of immoral, antisocial persons, such as gangsters and juvenile delinquents. Defenders of such communications point out that the behavior of such persons is labeled as bad. The present results indicate, however, that people may be influenced to become more like fictional persons who have undesirable characteristics, even when the irrationality of such qualities is clearly explained.

The present results also bespeak the power of all types of communication to effect at least superficial changes in the personalities of the recipients. The relative ease with which the reading of a single case history affected responses to F Scale items several weeks later suggests that stronger and more repeated communications can have more basic effects than the relatively superficial ones of the present study. In the present context, this result is disconcerting, since the characteristics the subjects acquired are thought by most social scientists to be undesirable. Of course, if the models presented have desirable qualities, e.g., are less authoritarian, then the potency of these types of communication may be socially valuable.

SUMMARY

Eighty-five subjects were first given the F Scale and a measure of anti-Negro prejudice in regular college classes. Four weeks later they read a case history of a person who was prejudiced and authoritarian. Half the subjects were told that the person in the case history was similar to themselves in background and objective personal characteristics. The other half were told that the person was dissimilar to them. Three weeks later the F Scale and the anti-Negro prejudice scale were readministered. It was found that those low in prejudice at the first administration became more prejudiced but only if they were told the case history was about someone like themselves. Persons low in the Defensiveness subscale of the F Scale also became more prejudiced by the third administration, if the person in the case history was similar to them in background. The latter group also became significantly more authoritarian, as measured by the Defensiveness subscale at the third administration. These results were interpreted in terms of the social support for suppressed

"wrong" feelings and in terms of the pressures to achieve cognitive consistency in one's self-concept.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

BERKOWITZ, L., & LUNDY, R. M. Personality characteristics related to susceptibility to influence by peers and authority figures. *J. Pers.*, 1957, **25**, 306–316.

KATZ, D., SARNOFF, J., & MCCLINTOCK, G. Ego-defense and attitude change. *Hum. Relat.*, 1956, **9**, 27–45.

MCCLINTOCK, C. Personality syndromes and attitude change. *J. Pers.*, 1958, **26**, 479–493.

STOTLAND, E., KATZ, D., & PATCHEN, M. The reduction of prejudice through the arousal of self-insight. *J. Pers.*, 1959, **27**, 507–531.

STOTLAND, E., ZANDER, A., & NATSOULAS, T. The generalization of interpersonal similarity. *J. abnorm. soc. Psychol.*, 1961, **62**, 250–256.

WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology.* New York: Holt, 1954.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 275–281

ON GROUP COMPOSITION

WILLIAM C. SCHUTZ
University of California, Berkeley

THE problem of composing groups to attain certain group outcomes has been of interest to social scientists especially in the last few years. Interest arises in part because prediction of group outcomes from a knowledge of the characteristics of the individuals and of the laws of interaction constitute an empirical test of the ancient problem of "emergents"—whether or not a whole can be predicted from a knowledge of its parts. Also, group composition presents an ultimate challenge to any theory of human interaction. If the theory is valid, it must predict certain outcomes when particular people interact. Finally, the problem of group composition has important practical implications for composing groups such as therapy groups, work teams, committees, research teams, hospital wards, army squads, and any other situation in which particular qualities are desired in particular groups.

After an extensive review of the literature on group composition, Gross (1959) concludes:

> In this investigator's opinion, there have been three major systematic attacks upon the group composition question: (a) Powdermaker and Frank's study on Veteran's Administration patients undergoing group therapy (1953); (b) the National Training Laboratory in Group Development studies on small group functioning (Stock and Thelen, 1958); and (c) Schutz's work on his *FIRO: A Three-Dimensional Theory of Interpersonal Behavior* (1958) (p. 7).

Powdermaker and Frank evolved three major group-relevant composition areas: the extent to which the patient is openly aggressive or competitive, the ability of the patient to tolerate aggression toward himself, and the patient's ability to expose his own weakness. The studies by Stock and Thelen (1958) conclude that important group composition variables are Bion's three modalities, pairing, dependence, and fight-flight.

The approach taken by the present investigator in past studies is based on the concept of compatibility. This approach assumes that wholes are predictable from a knowledge of the parts *and* the laws governing the interaction of the parts. In terms of three basic interpersonal needs—inclusion, control, and affection—various indices of compatibility are computed for pairs of people, based on the characteristic ways in which they express behavior in these areas and how much of each kind they want from others. From these considerations two kinds of compatibility are computed, one based on a notion of need complementarity called *reciprocal compatibility*, and one based on a notion of need similarity called *interchange compatibility*.

These two types of compatibility can be computed for each of the three need areas. Several experiments (Schutz, 1958) have been performed that show the technique to have a certain degree of validity for composing productive, cohesive groups desirous of further mutual interaction (e.g., roommates, traveling companions). One study used a derived technique to compose therapy goups (Gross, 1959).

The present study continues this interest in group composition, but differs in that (*a*) it deals primarily with variations of interchange compatibility, (*b*) it uses a different criterion from productivity, namely, awareness by the group members of the interactional characteristics of the group.

If groups are composed to be homogeneous with respect to the group members' expressed behavior and/or the behavior they want from others in the interpersonal areas of inclusion, control, and affection (as measured by FIRO-B), it is then hypothesized that after enough time has elapsed for the group members to become familiar with the group (six meetings in this case), the members should be able to identify correctly the interpersonal characteristics of their own group.

METHOD

Subjects and Setting

This study was performed on 70 subjects divided into five groups of 14 each, at the Western Training Laboratory in Human Relations in July 1959. The groups were heterogeneous with respect to sex, age,

education, occupation, and income. The laboratory is designed for: "... persons involved in educational, consultant, leadership or administrative capacities. School administrators, business executives, training directors, religious education workers, teachers, community organizations and civic leaders—in fact all those with major responsibilities for working with or through other people can profit from this laboratory." This statement (WTL brochure, 1959) describes well the actual distribution of participants.

The groups met once a day for 10 meetings of 1 hour and 45 minutes each, and the members had close contact with each other throughout the rest of the day, since the laboratory was located in an isolated residential setting. The meetings were run as unstructured groups with no agenda and with the aim of studying group process and interpersonal behavior through observation of the interaction among the group members. This type of group has been given various names: T group, training group, sensitivity training group, development group, basic group.

Procedure

Before any subjects met in groups, they were administered the FIRO-B (Fundamental Interpersonal Relations Orientation-Behavior) questionnaire. This instrument is composed of six Guttman scales of 10 items each, measuring an individual's expressed and wanted behavior in the interpersonal areas of inclusion, control, and affection (Schutz, 1958). On the basis of individual scores, the 70 participants were divided among five groups, each of which had members who were similar on certain of these characteristics. Following are the descriptions of the five groups based on measures derived from the FIRO-B. (For purposes of exposition the groups are designated by the letter of their FIRO-B descriptions. In the actual situation, of course, symbols unrelated to the descriptions were used to label the groups.)

Group A is characterized by the fact that the majority of group members prefer to initiate behavior of any kind more than they like to receive it. This includes initiating interaction rather than waiting to be invited, giving orders and directions rather than receiving them, and to a lesser extent giving affection rather than receiving it. In general the desire for interaction with people is relatively low in this group, while the concern with authority and control is very high.

Group B is characterized primarily by a concern for the area of control and authority. Like Group A, the members prefer to exert control rather than submit to it. They differ from A in that they prefer an average amount of interaction and have a considerably higher preference for receiving affection than initiating it.

Group C is characterized mainly by its strong preference for receiving behavior of all types rather than initiating it. This includes taking orders and directions rather than receiving them, receiving affection rather than giving it, and to a somewhat lesser extent being invited to interact rather than originating contact. The group members very much prefer to be interacting with people than not. Their desire to control other people is very low.

Group D is strongly and primarily concerned with affection, both giving and receiving, although their preference is to receive. In the area of control they prefer to receive direction rather than to give it. They are quite high in interacting and in their desire for interaction.

Group E is marked by a very low desire for close personal relations. The members also exhibit a strong preference for initiating interaction and contact with people rather than having others do it. They have a preference for receiving direction rather than giving it, though not as markedly as in Groups C and D. They are also moderately low on general interaction both initiating and wanting.

Prior to examining the data the experimenter told the subjects that Groups A and B were similar in terms of FIRO-B group description, and also that Group C was similar to Group D. Group E was not especially similar to any others. This classification was done to allow for later analysis of the groups in terms of three types (AB, CD, E) as well as five.

After six meetings all members of each group were given a dittoed sheet on which the group descriptions were written as above. Instructions were:

Below are descriptions of the five afternoon groups as predicted from FIRO-B scores. In order to test the validity of these composition patterns, we would like you to evaluate your own group. Read over all five descriptions. Then select the one that seems best to describe your afternoon group. Place a 1 next to that description. Then select the one next most like your group and place a 2 next to that one, and so forth, finally placing a 5 next to the description that sounds least like your group.

RESULTS

The original raw data[1] consist of rankings, from one to five, of the five groups in terms of respondents' perception of similarity of each description to his own group. These rankings were analyzed in two ways: by weighting all ranks and summing them, and by first choice rank only. The weighted ranking data will be considered first.

Percentage of Weighted Ranks

The ranks were weighted in inverse order; i.e., a ranking of 1 was given a weighting of 5, rank 2 a weight of 4, etc. For a given group, the sum of weighted ranks given to each of the five group descriptions by all 14 members of that group was computed. By calculating the percentage of the total score given to each of

[1] The author is indebted to Vernon Allen for data analysis and computation.

the five descriptions, the "percentage of weighted ranks" was obtained for each group. The hypothesis predicts that scores for Group A, for example, will reveal the highest percentage of weighted ranks for the Group A description, and so on for the other groups.

In order that the data be presented more meaningfully, these scores were transformed to take account of the limited range imposed by the ranking procedure. For example, when five categories are used, a description that happened to receive first place votes from all group members would have a weighted rank percentage of 33, while if the description received all last place votes it would have a weighted rank percentage of 7. The transformation was accomplished by setting 7% equal to 0 and 33% equal to 1. Then each obtained score is expressed as a proportion of the possible range of weighted rank percentages, a number between 0 and 1, with chance equal to .50.

In order to determine the error in prediction resulting from the high similarity of group descriptions among the possible choices, the proportions were computed for two situations: when the five descriptions were considered separately; and when the five descriptions were reduced to three by combining A with B, and C and D, as announced prior to the data analysis. The results of the transformed percentage of weighted ranks for both the five and three description situations are given in Table 1.

Using the *t* test, three groups, A, B, and D, were predicted significantly greater than chance, while the other two groups feli at about the chance level. Overall for the 70 individual predictions from all five groups the results were significant at the .01 level. Such an overall rate of predictive success (.70 to .75 correct when chance is .50) is typical of several studies using the FIRO composition technique (Schutz, 1958; Schutz & Gross, 1961).

First Choices

The second method of handling the results was to treat only the first choice given by the group members. The number of first choices given by all group members of a particular group was converted to a proportion of the total choices (14). Proportion of first choices as predicted is shown in Table 2, again with both five and three descriptive categories used for analysis.

TABLE 1

FOR WEIGHTED RANKS, PROPORTION OF PERFECT PREDICTION (1.00) ACHIEVED BY EACH GROUP (chance prediction .50)

Group	Using five descriptions	p	Using three descriptions	p
A	.88	$<.01$	.97	$<.05$
B	.86	.05	.88	.05
C	.54	—	.44	—
D	.77	.05	.91	.05
E	.54	—	.50	—
All groups combined	.72	.01	.74	.01

TABLE 2

PROPORTION OF GROUP MEMBERS CHOOSING THE CORRECT DESCRIPTION OF HIS GROUP AS FIRST CHOICE

Group	Using five descriptions	p	Using three descriptions	p
A	.57	$<.05$	1.00	$<.01$
B	.66	.05	.86	.05
C	.33	—	.53	—
D	.50	—	.92	.05
E	.06	—	.12	—
Chance	.20		.33	
All groups combined	.42	.01	.69	.001

For five descriptive categories, Groups A and B are again well predicted and Group D just misses significance (using the *t* test). When similar descriptions are combined a sharp improvement occurs, making D once again significant. Although the .05 level of significance is not attained, even Group C is predicted about 1.5 times chance for first choices. Every member of Group A predicted either Group description A or its similar description, B. As might be expected, the analysis by weighted ranks did not differ significantly when descriptive categories were reduced from five to three, but the analysis of first choices was significantly affected. This finding shows that there was a close association between the choice of adjacent, similar descriptions.

Summary of Results

Groups A, B, and D were predicted significantly better than chance by almost any measure, Group C showed a tendency to be well predicted as far as first choices were concerned, and Group E was poorly predicted by any measure. With all groups combined, the predictions were significantly higher than chance and lend considerable weight to the hypothesis and added confidence in the technique of group composition.

DISCUSSION

Why do Groups A, B, and D demonstrate such a high degree of predictive accuracy compared to Groups C and E? An answer seems to lie in the nature of the common characteristic by which each group was originally composed. Members of Group E had in common a desire *not* to do something, namely, not to have close personal relations. There was a great variety of preference for the behaviors they characteristically *do* express. Hence the behavior directly observable to group members was of a highly variable nature, a point reflected in their even distribution of choices on all five group descriptions. Within each of the well predicted groups, A, B, and D, there was a commonality of expressed behavior, either inclusion, control, or affection. The members of Group C were similar in that they all *wanted* someone else to initiate behavior. The commonality lay in the realm of want rather than of expression. Group C differs from E in that the members of E had a common wish to avoid something. The intermediate position of C with respect to predictability—not so good as A, B, and D, but not quite so bad as E—may be understandable in these terms.

The results suggest that groups composed around a common interpersonal characteristic are more identifiable by the group members after several meetings to the degree that the common trait is overtly *expressed*. If it is only a common *desire* rather than an overt expression on the part of the members, the common group characteristic is harder to detect, and if this characteristic is a behavior or feeling that everyone wishes to *avoid*, it is hardest of all to identify.

The relation between this finding and the Postulate of Group Development (Schutz, 1958) requires exploration. In part, the postulate states that every group goes through the same developmental stages in the same order: namely, concern first with problems of inclusion (in or out), then with problems in the control area (top or bottom), finally reaching the area of affection (near or far). In the present study there was no direct test of the postulate; it was simply noted that one area was stressed in each group. The other two areas may have been dealt with briefly before each group focused on its salient area, but there are no relevant data on this point.

Behavioral Observations

Although these groups were composed on the basis of homogeneity along the inclusion, control, and affection scales of FIRO-B, no specific predictions were made about the behavior of the groups. This was because the interchange compatibility (Schutz, 1958) that was built into every group did not lead to any clear behavioral prediction about the content of the topics and the central concerns of an unstructured group. However, informal impressions of group behavior revealed striking differences among the groups, differences that seemed generally consistent with the composition patterns.

The group trainers[3] were asked, about 4 months after the workshop, to describe the behavior of their groups. Since they were at that time familiar with the experiment they cannot be considered independent observers. However, they attempted to be as accurate and as objective as possible, and their reports are helpful for illuminating the behavioral phenomena corresponding to the objective FIRO-B composition patterns and are suggestive for understanding, future hypotheses, and applications. It is in this sense that they are presented, rather than as experimental evidence.

The general characteristic of the best predicted four of the five composed groups was

[3] The trainers for Groups A, B, C, D, and E were, respectively, Schutz, John Weir, Robert Tannenbaum, Harrington Ingham, and Joseph Luft. The valuable cooperation of the last four is greatly appreciated.

a single focus of discussion, especially at the beginning meetings. In heterogeneous groups a wide variety of topics is typically discussed throughout the life of the group. In contrast, in each of these four composed groups there was a narrowing of discussion topics to the point where there was predominantly one area at issue.

In the two initiating, controlling groups (A and B) the discussions centered around authority, power, and competition:

In Group A the atmosphere for most of the first several meetings was one of tense caution and covert struggle. Each member seemed afraid to attempt any leadership act because of the perceived strength and power in many other members, which might be turned against the initiator. At one point a group member described the atmosphere thus: "I feel like there's an electric charge circling through the group and each time it comes to me the hair on the back of my neck stands up."

There was a great deal of talk about the members' relation to their boss and their relations with the trainer of the group. During the hours between meetings, for example, various members of Group A challenged the trainer to a ping-pong match, "jokingly" punched him on the arm, and tried almost to the point of personal injury to hit a softball farther than he did. In addition, one group meeting was almost disbanded so that the group could go out on the grass to see whether a particular group member could throw a football farther than the trainer. There were many additional manifestations of concern with masculinity. The group members soon named their group the "Tigers," later changed to the "Gorillas." In the morning meetings there was a great deal of talk about hangovers and the previous night's drinking, and the group members were highly overrepresented each night among the people who, after the center had closed, stayed up to continue drinking and socializing.

After several meetings this common group interest became clearly established. Attempts to understand the dynamics of competition and authority problems in group behavior were initiated when one group member said he felt "inadequate" in his job, and the people who promoted him did not realize his inadequacy. From this point on the group worked through from job inadequacy to social inadequacy to a felt inadequacy with women (the group had 12 men and 3 women). Then ensued discussion about mothers and fathers who never gave reassurance, or permitted achievement and the attaining of competence, as possible explanations of their inadequacy feelings and hence their attitudes toward competition, authority, power, and masculinity.

Hence in the short space of 10 meetings in which usual groups of this kind cover more topics to less depth, this group concentrated on one problem area—authority and competence—and carried it through to a rather remarkable depth of insight.

Group B, made up of people with a strong preference for controlling others, according to FIRO-B, is described as follows by the group trainer:

The composition of the group was obvious to trainer and members almost from the beginning. The initial meeting was devoted to a heated denunciation of the behavior of a lab delegate who had appointed himself leader in a somewhat highhanded manner during a preceding general session. The group was very hostile toward him, resenting his leadership as presumptuous and censuring him as an authoritarian.

The first four or five group meetings were devoted to competitive leadership strivings. The group completely rejected any attempt by the trainer either to impose or insinuate training procedures or structure. There was much hostility directed toward him and at one time a majority decision to eject him from the group. This leadership attempt met with strong resistance from a vocal minority, resulting in the decision's being changed to one of letting the trainer sit with the group, but ignoring him.

At the same time the group members were struggling to establish warm affectionate relationships with each other. This need was finally brought into the open during the last five meetings and, when mutually shared and experienced, brought a diminution in the competition for leadership and a marked group integration. However, when the lab design later required each group to observe and be observed by another group the rebellion toward authority flared up again almost to the point of an outright refusal to cooperate with the plan. Ultimately the group acquiesced to the observation procedure. During the observation of another group, the need to control became so intense one member said, "Aren't you (the other group) going to ask us what we think of your meeting?" Finally the drive to exercise control could no longer be contained. The members burst out with observations, evaluations, and reactions to what they had observed, alternately praising their own group behavior and criticizing the behavior of the other group. Analysis and discussion of this even developed much insight into the dynamics of the earlier group behavior regarding authority and control.

The final meetings were characterized by the exchange of affection and mutual acceptance. Conflict over authority, control, and structure disappeared and it became possible to explore the dynamics of authority and affection in a manner that was very deep and satisfying to the members. Several testified to the gaining of important new insights into their relationship with authority figures.

Thus in this group, too, the predominant theme was control and authority, with affection relations beginning to be evidenced toward the end. It is interesting to note that one member of this group was at variance with

the others from the outset. When they were so preoccupied with control she wanted to talk about expressing feelings and warmth. She could not believe she was like the other members (according to the composition pattern). Often she wrote of peace and conciliation on postmeeting reaction sheets distributed after each meeting, when virtually all the others would comment on strife and conflict. It turned out later that she was placed in the group only because of the mechanical limitations to making every group homogeneous. Actually, according to her FIRO-B score, she was very low on expressing control (2 on a scale of 0–9) compared to the others, and considerably higher on affection. She was quite relieved to discover this.

Group D, composed of people high on the affection scales of FIRO-B, was described by its trainer in the following way:

> There was exceedlingly little discussion of authority or power, but a great deal about expressing and accepting feelings. In the beginning there was much talk of getting along fine in the group because everyone was considerate and understanding. Then there was an uncomfortable time when hostility was condemned for a surprising number of reasons, followed by a painful realization that the people could not really communicate without including the hostile things they felt. Even after this conclusion was accepted it remained unwelcome.
>
> Throughout the life of the group people could speak of their feelings of inadequacy. It was no problem for the group to accept and encourage this. Several members considered increasing prominence as a questionable thing. They recognized their desire for it, but also considered that it detracts from family life, and may distort a person's values about the work itself.

Here the emphasis was almost entirely on affection and expressions of emotion, with almost no attention to the problems that preoccupied Groups A and B. It is interesting to note that discussion of inadequacy, so difficult to face in high-initiating Group A, was easily done by Group D.

Group C, made up of participants with a preference for receiving behavior in all areas, especially control and affection, rather than initiating it, is described this way by its group trainer:

> I certainly had no difficulty in getting the group to follow most of my suggestions (there was really only one man who would resist and fight back). Actually the group gave very little overt attention to the control area. The chief characteristic of the group was its very heavy occupation during most of its meetings with feelings. I believe the group members were aware of many more feelings than they were able to express. Anxiety or other concerns seemed to be associated with the expression of strong positive or negative feelings. Most members were unable to use themselves in order to facilitate the expression of feelings by others. Many had difficulty in expressing great discomfort in facing these strong feelings of others. The group had one of its most intensive experiences on 2 days in which all members of the group functioned with eyes closed. This seemed to make it easier to express, facilitate in others the expression of, and to receive the expression of strong feelings. Certainly there was great need to receive affection but a considerable inability of group members to express affection. To a lesser extent, this would be true with inclusion. No one really expressed strong control and, as previously stated, the members were quite receptive to control.

Here the chief characteristic of the group is its difficulty in expressing feelings. The anxiety they feel in initiating behavior and their need to receive affection and inclusion, and great willingness to accept control, would be consistent with the composition pattern. Also consistent with the prediction is one contrast between Group B—the high controllers—and this group. In Group C the trainer had no difficulty in getting the group to follow his suggestions, while in Group B the members completely rejected any such attempts by the trainer and actually once voted to eject him from the group.

Group E, composed of persons described on FIRO-B as low on affection, was the poorly predicted group. Perhaps a description of the behavior of this group by the trainer will shed light on the predictive failure:

> There were two major factions, roughly authoritarian and equalitarian (or high controllers and low controllers). Although the low controllers did not initiate actions they were highly effective in stopping the more agressive controllers. The low control people actually turned out to be stronger and eventually set the tone and norms of the group. The opposing factions locked horns early and for a day and a half the group was stalemated. This was symbolized by the copper pot which was used to collect anonymous statements regarding some significant events as seen by the group members. For nearly 2 days the group could not resolve the simple matter of extracting the slips of paper from the pot. Every time the controllers initiated action, the others systematically exposed their motives and refused to agree to any action which forced even one person to do anything against his wishes. They finally found a way out of their bind, with the result

that everyone had lived through an experience of clarifying and resolving strong differences. This experience, followed by others like it but of lessened duration, was heuristic, to say the least. Even the controllers were surprised at the affection they felt and could give as the crises were overcome. Having tamed tigers in their midst, they felt, when they visited the "Tiger" group, that the latter had a long way to go.

The problem of affection did not arise until late in the group in the form of surprise at the affection they felt. Perhaps the struggles of the group at the beginning were actually methods for maintaining emotional distance among the group members.

Thus if the FIRO-B composition patterns are accepted as accurate, the interesting result of the composition patterns was the alteration of the topics discussed to bring them into line with the "common group problem." The possibility thus emerges of composing groups with certain characteristics for a given purpose. For example, a training group to be introduced to a variety of group phenomena might better be made heterogeneous on the FIRO-B scales, while a therapy or other type of group in which it is desirable to explore a single area in some detail could be composed homogeneously.

SUMMARY

As part of the Western Training Laboratory in Human Relations, five 14-person groups were composed of persons with certain similar individual interpersonal needs, as determined by scores on the FIRO-B questionnaire. After meeting six times the group members were given behavioral descriptions of the five groups with instructions to rate the five descriptions in terms of similarity to their own group.

Three of the groups were accurate, identifying their own group descriptions significantly better than chance. One group showed a tendency to select the "correct" description for first choice, and one group was poorly predicted. The total prediction was highly significant and lends confidence to the FIRO technique for group composition.

An explanation offered for the results was that homogeneous groups are predictable to the degree that the common characteristic is *expressed* behavior, rather than a common *desire* or a common wish to *avoid* a particular interpersonal area.

Behavioral differences in the composed groups were noted. Each group seemed to settle on a particular topic, and went into this topic to a greater depth than is usual in groups of this type. Practical implications were noted.

REFERENCES

GROSS, R. L. Therapy group composition: Personal-interpersonal variables. Unpublished doctoral dissertation, University of Utah, 1959.

SCHUTZ, W. C. *FIRO: A three-dimensional theory of interpersonal behavior.* New York: Rinehart, 1958.

SCHUTZ, W. C., & GROSS, E. F. The FIRO theory of interpersonal behavior: Empirical tests and applications to business administration. In, *Contributions to scientific research in management.* Berkeley: Univer. California Press, 1961.

STOCK, DOROTHY, & THELEN, H. *Emotional dynamics and group culture.* National Training Laboratories, 1958.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 282-287

AN INVESTIGATION OF UNILATERAL CEREBRAL PATHOLOGY WITH EQUATED VERBAL AND VISUAL-SPATIAL TASKS[1]

ROSEMARY STARK[2]
Wayne State University

THE relationship of unilateral cerebral pathology to selective psychological deficit has been the subject of increasing experimental investigation over the past decade. Several studies (Anderson, 1951; Heilbrun, 1956; Reitan, 1955) have lent confirmation to the well known clinical observation that verbal deficit is more strongly associated with left hemisphere pathology and have further suggested that performance deficit may be more strongly associated with right hemisphere pathology. The verbal performance dichotomy was based primarily on the sums of the respective Wechsler-Bellevue Intelligence Scale scores. Examination of the consistent findings in these studies suggested that the performance deficit in subjects with right hemisphere pathology may be more specifically described as a visual-spatial deficit. Further, it was felt that tasks developed to maximize cerebral pathology would serve better than tests of general intelligence to measure such deficits. The present study was designed to investigate these relationships with paired-associate learning tasks in the verbal- and visual-spatial modalities.

Paired-associate learning tasks were selected because they are well suited to maximize the tendency toward perseveration, a characteristic often observed in the performance of brain injured subjects. By perseveration is generally meant the persistent, nonadaptive repetition of an initial response through changing stimulus conditions. As such, it was observed in brain injured subjects by both Benton (1955) and Bender (1938) in their respective visual-spatial tests. It may be thought of as resembling the tendency to "repetition" noted by Piotrowski (1937) on Rorschach test protocols or as stemming from the "stimulus bondage" of brain injured subjects described so well by Goldstein and Scheerer (1941). The latter authors stressed particularly the tendency of brain injured subjects to fixate one aspect of the stimulus situation and their inability to shift to others.

For the purposes of this study, perseveration is defined as the persistent repetition of responses of high associative strength where they are no longer appropriate. The paired-associate tasks were structured so that they would first elicit and reinforce a strong association to a given stimulus and then require the learning of a new, unrelated response to the same stimulus. Such a procedure should maximize the tendency to perseverate the strong association to the stimulus throughout the learning of the new response to the same stimulus.

It was hypothesized that:

1. On the verbal task, subjects with left hemisphere lesions show poorer verbal scores and more perseverative errors than either the right hemisphere lesion group or a control group with no cerebral pathology.

2. On the visual-spatial task, subjects with right hemisphere lesions show poorer visual-spatial scores and more perseverative errors than either the left hemisphere group or the control group.

METHOD

Tasks. The stimuli and responses for the verbal task are shown in Table 1. The seven stimulus words were chosen at random from Shellenberg's frequency table (Tinker, 1947). Since frequency of response has been interpreted as indicating strength of association, those associative responses most frequently given were chosen as the strong or "easy" response. Seven other words, chosen from the same list because of their apparent unrelatedness to the stimulus words, served as the "difficult" competing response.

The visual-spatial task was constructed in the same way. Figure 1 gives a sample of the items. Seven such simple, incomplete stimulus figures were selected for their apparent property of eliciting simple "closure" responses.

Both tasks were tried out on a normal population

[1] Condensed from a doctoral dissertation submitted to Wayne State University, completed under the direction of Bertram D. Cohen with the assistance of Michael M. Reece.

The author wishes to express appreciation to John E. Webster, MD and Stephen E. Gurdjian, MD for their cooperation in the referral of patients with thorough diagnostic evaluations.

[2] Now at the University of Detroit.

TABLE 1

STIMULI AND RESPONSES FOR THE VERBAL TASK

Stimulus	Easy response	Difficult response
DARK	light	salt
SLOW	fast	needle
CARPET	rug	whistle
HIGH	low	child
LONG	short	health
BED	sleep	street
BLOSSOM	flower	hammer

with the general aim of selecting responses that would give equivalent error scores. It was found that normal subjects did respond to most of the verbal stimulus words with the selected easy response and to the visual-spatial stimulus figures with the simple closure response. Those difficult visual-spatial responses were selected which yielded a mean error score equivalent to that on the verbal task.

Subjects. Three groups of 20 subjects[a] each composed the sample in the final study. All were private patients of Detroit neurosurgeons, hospitalized for recent onset or recurrence of illness as follows: a control group of tack injury patients (C group) in whom cerebral pathology was ruled out, a right hemisphere lesion group (R group), and a left hemisphere lesion group (L group). All groups were screened for background factors of serious mental or emotional disturbance. All had been functioning effectively in their normal occupations prior to the onset of illness. None had been institutionalized other than for brief periods of private medical hospitalization. Only right-handed subjects were included. Current functioning level was determined on the basis of verbal IQ estimated from the Vocabulary subtest of the Wechsler Adult Intelligence Scale. A minimum verbal IQ of 80 was accepted; all but eight of the subjects scored 90 or above. All groups were equated for age and vocabulary scaled scores as shown in Table 2, and the two experimental groups were well matched in educational background.

The brain injured subjects were diagnosed by a neurosurgeon following a complete neurological study including a neurological examination, an angiogram, an electroencephalogram, and, in some cases, a ventriculogram, pneumoencephalogram, or skull X ray. These subjects were of two kinds, recent cerebrovascular accidents and brain tumors.

Since these were all private patients, only those who consented to take part were tested. Each subject was tested individually during his hospitalization by the same examiner.

Procedure. Since the aim of the easy task was to reinforce the strong association and promote perseveration of this response in the learning of the difficult list, the easy list was presented to the criterion of two perfect trials or a maximum of three trials, this number being found sufficient for perfect learning in the normal population. On each trial of the tasks, each stimulus was presented in succession with a timed interval for the subject to respond before presentation of the correct response. Five seconds were allowed for each verbal item and 10 seconds for each visual-spatial item, the longer time for the latter necessitated by the instruction to draw the figure completion. The items were randomized for each trial. The first trial of the difficult list on each task was made a guessing trial to maximize perseveration tendencies. On this trial the subject was instructed to give any new response that came to mind as long as it was different from that on the easy list. The difficult list was then presented to the criterion of two perfect trials or a maximum of six more trials. Possible fatigue or practice effects were controlled by random assignment of all subjects to two order conditions. Half the subjects were given the verbal task first and the other half the visual-spatial task first.

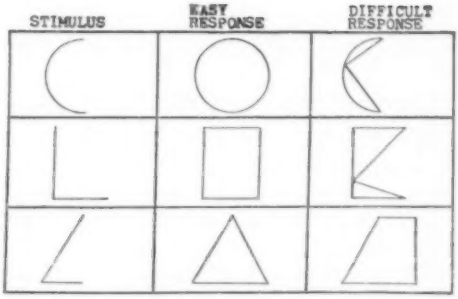

FIG. 1. Sample stimulus figures and responses for the visual-spatial task.

Scores. Only the responses from the difficult list were scored. The error score consisted of all the incorrect or missing responses from the last six trials of the difficult list. The first trial could not be counted in the error score since the subject had no way of anticipating the correct response. The perseveration score included all seven trials of the difficult list and consisted of all the responses which were repetitions of responses from the easy list.

RESULTS

Error scores. An analysis of variance technique using a mixed design considering groups, tasks, and order on the error scores for the three groups, divided according to the order conditions, was performed first. Since no significant effect of order was indicated, the order conditions were combined. The means and standard deviations on the error scores for the three groups are seen in Table 3. Inspection of these means shows that the tasks were of approximately equal difficulty for the C group,

[a] Complete information relevant to subject population, tasks, and statistical treatment of the data can be obtained from University Microfilms, Ann Arbor, Michigan, Doctoral Dissertation Series. The dissertation title is the same as that used for this article.

TABLE 2
DESCRIPTION OF SUBJECTS

Group	Mean Age	Mean Years Education	Lesion			
			Kind		Duration	
			CVA	Neoplasm	3 months	3 months+
C	49.7	9.8				
R	51.5	11.8	18	2	11	9
L	50.6	11.2	17	3	11	9

TABLE 3
MEAN ERROR SCORE

Task	Groups					
	C		R		L	
	M	*SD*	*M*	*SD*	*M*	*SD*
Verbal	19.80	7.14	19.05	9.06	30.95	8.07
Visual-spatial	21.35	7.16	32.55	8.43	23.30	8.43

but selectively difficult in the predicted direction for the two brain injured groups. While the R group scored the same as the C group on the verbal task, they scored on the average approximately 11 more errors on the visual-spatial task. The L group, on the other hand, did almost as well as the C group on the visual-spatial task but scored on the average approximately 11 more errors on the verbal task. These mean differences between the experimental groups were found to be significant by t test ($p < .01$), whereas none of the mean differences between experimental and control groups reached significance.

These results are illustrated most clearly in Figure 2 where the means of the C group on the two tasks are used as the base line against which to compare the task means of the brain injured groups. The hypotheses concerning the error scores for all groups were confirmed.

Perseveration scores. The results for the perseveration scores were not as clear-cut as for the error scores. Since a preliminary analysis showed no significant effect of order of administration, the means for both order conditions were combined. These means are shown in Table 4. Although an analysis of variance yielded a significant F ($p < .01$), an inspection of the means suggested that this was due to

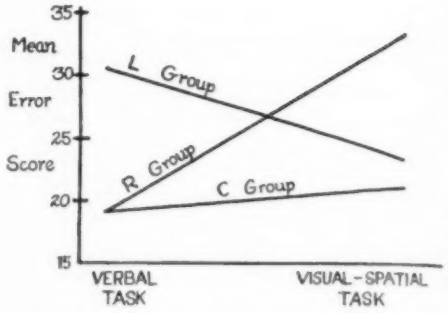

FIG. 2. Comparison of mean error score of all groups, the control group (C group) mean used as the base line.

the extreme perseveration of the R group on the visual-spatial scores. Here it is seen that this group scored an average of 14.70 perseverative errors in comparison with the C group average of 3.35 and the L group average of 5.90. These differences were both significant by t test ($p < .01$).

Differences between the means for the three groups on the verbal perseveration scores did not reach significance. It was felt, however, that the trend of these means from 1.70 errors in the C group to 3.60 errors in the L group did indicate the presence of the verbal perseverative tendency in the latter.

Diagnostic value of the tasks. An analysis of the distribution of individual scores was made to assess the degree to which the learning task scores could identify an individual diagnostically. Cutting scores were determined by graphing the distributions of the three groups separately for each task and selecting the highest crossover point which identified the brain injured group in question. On the verbal task, 85% of the C and the R groups scored below the cutting score, while 70% of the L group scored above it. On the visual-spatial task, 90% of the C group and 85% of the L group scored below the cutting score, while 75% of the R group scored above it.

The percentages of correct identification for each learning task considered separately, however, cannot be considered evidence of diagnostic accuracy. A subject must not only be identified as showing evidence of deficit on the appropriate task; this deficit must also clearly

TABLE 4
MEAN PERSEVERATION SCORE

Task	Groups					
	C		R		L	
	M	*SD*	*M*	*SD*	*M*	*SD*
Verbal	1.7	1.54	3.4	2.59	3.6	3.32
Visual-spatial	3.4	2.45	14.7	8.49	5.9	4.13

exceed the scores on the opposite task. When the evidence from both tasks was thus combined for each group, 75% of the C group and 80% of the R group were correctly identified. The accuracy in the L group, however, dropped to 55%. This drop was due to the relative task scores of three subjects. Although these subjects exceeded the cutting score on the verbal task, their visual-spatial error scores were also so elevated that they would be misidentified as right lesion subjects.

An examination of subjects misidentified in the two brain injured groups suggested that the factor of age may have a selective effect on performance on the two tasks. As a group, the mean age of these misidentified subjects was 57.5 as compared with a mean age of 51.3 for the entire brain injured sample. This discrepancy led to an investigation of the relationship between age and learning task scores in all groups. Pearson product-moment correlations were performed between age and each task error score. A correlation of .47 ($p < .05$) was found between visual-spatial error score and age; a correlation of .23 ($p < .10$) was found between verbal error score and age.

DISCUSSION

At an empirical level, the results of the present study suggest some clear relationships regarding localization of cerebral functioning. The differences on the error scores for both tasks make it possible to conclude that:

1. Subjects with left hemisphere pathology, although able to learn and retain visual-spatial designs almost as well as the controls, are significantly inferior to either subjects with right hemisphere pathology or controls in verbal learning and retention.

2. Subjects with right hemisphere pathology, although able to learn and retain verbally presented materials as well as controls, are significantly inferior to either subjects with left hemisphere pathology or controls in visual-spatial learning and retention. It should be noted that all subjects tested were right-handed; hence, the visual-spatial deficit in the subjects with right hemisphere pathology does not represent a motor deficit.

Although the results on the perseveration scores were not as clear-cut, they were in the predicted direction. They clearly demonstrate the prevalence of the perseverative tendency in both brain injured groups as compared with the controls. The fact that this tendency is better demonstrated on the visual-spatial task may be due to a difference in the degree of conscious verbal control exercised in the respective tasks. The more frequent response of the left hemisphere lesion subjects on the verbal task was blocking, i.e., no response. Subjects frequently commented that the response they "had in mind" was the old one from the easy list but they knew it was wrong. It may be that verbal responses are more subject to verbal control than are visual-spatial responses. Hence, the overt perseveration would be less likely to be demonstrated on verbal tasks than on visual-spatial tasks.

While the results obtained with these learning tasks show clear group differences, much individual variability remains. Although the cutting scores showed an average of about 85% accuracy for each task considered separately, this percentage dropped considerably when overlap on the task scores was considered. It was further evident that this percentage of success was due primarily to the right hemisphere lesion group where 80% of the subjects were correctly identified. Considering the left hemisphere group alone, it is clear that 55% accuracy is hardly sufficient to place much faith in the tasks for individual diagnostic purposes.

However, examination of the distributions of the two task scores revealed something that may be relevant in other psychological diagnostic tests in the area of cerebral pathology. This concerns the overlap of the visual-spatial scores. It was noted that the reason for the poor diagnostic accuracy in the left hemisphere

lesion group was not the failure to demonstrate deficit on the verbal task but, rather, the fact that older subjects with cerebral lesions tended to show elevated visual-spatial scores regardless of area of lesion. The significant positive correlation of age with visual-spatial scores in all groups further suggested that an increase of visual-spatial errors with age may also be expected in a normal population. These findings suggest that visual-spatial functioning is more easily disrupted by both age and cerebral pathology than verbal functioning. Hence, visual-spatial tests of cerebral pathology should be carefully equated for age if they are to have any validity.

An examination of individual test scores raised another point that may be important for both localization studies and the use of tests for diagnosis. Much of the variance on the respective tasks was found to be due to the extreme scores of several highly deviant individuals in both brain injured groups, e.g., a radiologist, a minister, an engineer, and a psychiatrist. Such subjects—individuals with good educational background and a general intactness in most areas of intellectual functioning—were found to be the best subjects for demonstrating clear localization relationships.

In summary, the tasks here developed, while showing very clear results on a group basis, at the same time show limitations for individual diagnostic use. The validity of such tasks could be greatly improved with norms for age and general level of functioning.

The implications of the present study for localization of cerebral functioning could be clarified in two needed follow-up studies. It is possible that the present lateral hemispheric division of brain injured subjects obscured important anterior-posterior relationships. It may be that the verbal-symbolic functions are more closely related to anterior portions of the brain and visual-spatial to posterior, as indicated, for example, by Benton (1955, p. 36). A definitive study would include at least four groups of experimental subjects: right and left hemisphere anterior, and right and left hemisphere posterior. Further, it is not known what possible effects cerebral dominance may have on this apparent functional asymmetry of the brain. The present study should be repeated on a left-handed population to clarify the importance of dominance in relation to laterality of lesion.

SUMMARY

This study was designed to investigate the relationships between left hemisphere pathology and verbal deficit and right hemisphere pathology and visual-spatial deficit. Equated paired-associate learning tasks were administered to three groups of 20 subjects each: a control group of back injured subjects in whom cerebral pathology was ruled out and two brain injured groups with demonstrable right and left hemisphere lesions.

The results confirmed the hypotheses that: subjects with left hemisphere pathology are significantly inferior to both subjects with right hemisphere pathology and controls in verbal learning and retention, and subjects with right hemisphere pathology are significantly inferior to both subjects with left hemisphere pathology and controls in visual-spatial learning and retention. Both brain injured groups performed as well as the controls on the tasks apparently unrelated to the area of their lesions.

A significant relationship between visual-spatial performance and age was demonstrated in all groups. The results suggested that visual-spatial functioning is more easily disrupted by both age and cerebral pathology than verbal functioning.

These findings suggest an apparent functional asymmetry of the brain in respect to verbal and visual-spatial tasks. Two follow-up studies are indicated: one in which the groups would be further subdivided within each hemisphere as anterior and posterior; the other, a repetition of the present study on a left-handed population to determine the possible effects of cerebral dominance on laterality of lesion.

REFERENCES

ANDERSON, A. L. The effect of laterality of localization of brain damage on Wechsler-Bellevue indices of deterioration. *J. clin. Psychol.*, 1951, 7, 149–153.

BENDER, L. A visual motor gestalt test and its clinical use. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1938. No. 3.

BENTON, A. L. *The revised visual retention test: Clinical*

and experimental applications. Iowa City: Iowa State Univer. Press, 1955.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior, an experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

HEILBRUN, A. F., JR. Psychological test performance as a function of lateral localization of cerebral lesion. *J. comp. physiol. Psychol.*, 1956, **49**, 10–14.

PIOTROWSKI, Z. The Rorschach inkblot method in organic disturbance of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525–537.

REITAN, R. M. Certain differential effects of left and right cerebral lesions in human adults. *J. comp. physiol. Psychol.*, 1955, **48**, 474–477.

TINKER, M. A. *Introduction to methods in experimental psychology*. (2nd ed.) New York. Appleton-Century, 1947.

(Received February 4, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 288-293

SOME DETERMINANTS OF SELF-EVALUATION

HAROLD B. GERARD[1]

Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey

THERE appear to be two types of social comparison that affect self-evaluation. In one process a person directly compares his standing on a given attribute with that of others; in the other, his self-evaluation is influenced by his conception of how others regard him, in a way reminiscent of Cooley's (1909) "looking glass" self. The distinction parallels one drawn by Deutsch and Gerard (1955) between informational and normative influence upon an individual's judgment regarding a nonpersonal object. Informational influence involves reliance upon another as an impersonal source of information concerning some object that is being judged, just as one would refer to an arbitrary measuring instrument. Normative influence can occur only when members of the comparison group are able to reward or punish the person for his acts and beliefs. When sanctions can be applied the individual tends to conform to the others' expectations of him. Direct and reflected comparison correspond to informational and normative influence, with the self as object.

Previous research (Festinger & Thibaut, 1951; Gerard, 1953) supports the expectation, for either type of social comparison, that the greater the discrepancy between the individual's current estimate and the estimate derived from comparison, the greater his tendency to change his current estimate.

The relevance of the distinction between the processes of direct and reflected comparison should be established if factors can be found that produce differential effects on self-evaluation in one type of comparison but not in the other, effects that can therefore be attributed to mediation by one rather than the other process. Any factor that affects the degree to which members of the comparison group have means-control over the person's satisfactions should tend to produce reflected comparison. Naturally-given individual differences in dependency upon others for self-evaluation should presumably also affect the degree to which the person adjusts to a discrepancy in a situation engendering reflected comparison. Logically, either experimentally induced dependency or a naturally given variable may serve the role of detecting reflected comparison. From the foregoing discussion these predictions follow:

1. Given a discrepancy between anticipated and actual standing in a group, change in self-appraisal should result; the greater the discrepancy, the greater the tendency to change.

2. If the skill involved is relevant to the individual's standing in the group, anticipated confrontation by others in the group should, in addition to direct comparison, produce reflected comparison. Reflected comparison in turn should result in an increased tendency for the person to change his self-evaluation in the face of a discrepancy between anticipated and actual standing.

3. The greater the person's dependence upon others for self-evaluation, the greater should be the tendency for reflected comparison.

METHOD

Subjects

A total of 159 undergraduates at New York University served as subjects. Fifty-one potential subjects had to be eliminated since they did not participate in all three phases of the experiment.

General Setting

The experiment was conducted in the classroom and used time during each of three class sessions. During the first session a specially devised questionnaire was administered that provided an a priori measure of dependency upon another for self-evaluation. Each item could be answered "true" or "false." Some items referred to personal characteristics whereas others confronted the subject with a hypothetical social situation and asked him how he would react to it. A large pool of likely items were collected from a number of sources and each item was thoroughly discussed with the author's colleagues in terms of its relationship to dependency upon others for self-evaluation. A tentative list of 71 items was administered to a comparable sample of subjects at another university.

[1] The author wishes to thank G. S. Rotter for help with the analysis of the data and M. Deutsch and W. M. Evan for critical suggestions regarding the manuscript.

Items that failed to discriminate, i.e., were answered by nearly everyone as either true or false, were eliminated, as were items that correlated poorly with the subject's total score. The final list upon which the present analysis is based consists of 27 items. Examples of items that refer to a personal characteristic are:

One of my secret dreads is to stand out like a sore thumb.

I don't mind being the butt of a joke.

Examples of items that confront the subject with a hypothetical situation are:

When someone asks me to lend him something, I usually can't refuse even when my better judgment tells me to refuse.

I feel uncomfortable when I enter a room where everyone has already assembled and is in a discussion.

To have induced differential dependency experimentally would have required closer control than can usually be achieved in a classroom situation; we therefore settled for natural variation in a tendency that could serve alternatively as a detector of reflected comparison.

During the second session, the Psychological Corporation's Space Relations test was administered. In order to heighten concern with performance, subjects were told that the ability involved is rather critical in most walks of life. The test items are quite difficult and it was presumed that the subject would be unable to predict with much certainty what score he had made. In other words, his expected score and self-evaluation should tend to be somewhat labile. Immediately after the test, a questionnaire was given in which the subject was asked to indicate the score he expected to make (the scoring system was carefully explained) and he was asked to guess the score he thought a typical member of a list of groups would have made. Among others, this list included a "group of your friends," "honor students at New York University," and "the lower 10% of students at New York University." The subject was asked for these estimates to provide a basis for determining his own relative level of aspiration. Since the test was introduced as a fairly good predictor of academic success, the subject's guesses as to how the best and poorest student would have fared were thought to establish a range within which his own score expectation could be evaluated. The questionnaire contained a six-point a priori scale on which the subject was asked to indicate his ability. He was also asked to indicate the score he would have made under ideal conditions. A number of other peripheral questions were included, to which reference will be made later.

The third session followed the second by 2 days. During that class period each subject received, on a slip of paper, the score he had presumably made. Another questionnaire was then administered which was substantially the same as the one administered immediately after the Space Relations test. Some of the wording had to be changed to accord with the fact that the subject now had his score.

The dependency questionnaire administration preceded the second session by 2 weeks and was so presented that the subjects could not connect it with what followed. Its eventual use, any other deception involved, and the general purpose of the experiment were revealed at the end of the third session.

Experimental Manipulations

In order to induce differential reflected comparison, approximately half of the subjects were led to believe that everyone's performance would shortly be public knowledge. They were told that a roster of scores would be passed around the class and then posted so that each person could see just where he stood in relation to particular people in the class. He would thus be confronted by the evaluations of others by virtue of their knowledge of his relative standing, with attendant, anticipated social consequences. The other half of the subjects were told that each individual's score was his own business and would under no circumstances be made public.

To vary discrepancy, the score given each subject was falsified such that within the public and private treatments a variation in the discrepancy between expectation and performance was produced. The subject received either a false low score or a high score. The false low scores were either 10, 11, or 12 whereas the false high scores were 67, 68, or 69. A presumably accurate but actually false distribution of scores was written on the blackboard, and it was made clear that the high false scores placed the individual in the upper 5% of the class, whereas the scores selected as the low false scores placed the individual in the lower 5% of the class.

RESULTS

Changes in Self-Evaluation and Level of Aspiration

An index of expected performance was derived from the questionnaire administered immediately after the test. The index, I_e, was computed as follows:

$$I_e = \frac{E_e - E_{10}}{E_H - E_{10}},$$

where E_e = subject's expected score, E_{10} = subject's guess as to the score a member of the lower 10% of the class would make, and E_H = subject's guess as to the score an honor student would make. The index represents the subject's estimate of the quality of his own performance relative to his estimate of the general range of scores for the student body at large. Subjects with an index at or above the median were classified as expecting a high score, whereas subjects with an index below the median were classified as expecting a low score. The two degrees of score expectation were cross-classified with the high

and low false score feedback to yield four treatments, two in which the discrepancy between expectation and presumed performance was relatively small and two in which the discrepancy was relatively large.

The data presented in Table 1 indicate the mean shift in self-evaluation by subjects in all 16 experimental conditions. If we examine the amount of change produced by the three experimental variables—public vs. private (Cells 1–8 vs. Cells 9–16), high discrepancy (where presumed and expected performance were quite different, i.e., Cells 3–6 and Cells 11–14) vs. low discrepancy (where presumed and expected performance tended to be somewhat alike, i.e., Cells 1, 2, 7, 8, 9, 10, 15, and 16), and high vs. low dependent (odd vs. even numbered cells)—we find that, overall, each of the variables produces significant effects in the expected direction. However, it is apparent that the interactions between the experimental variables are such that it is not meaningful to consider their effects independently of one another.

A change in self-evaluation should be reflected in a change in level of aspiration. On the questionnaire administered immediately after the test subjects were asked to indicate what score they would have received under ideal conditions (with no distractions, etc.). The same question was asked after they received their scores. Table 2 presents the mean change in the score that would have been made under ideal conditions, from before to after the presumed score was received. These results also reflect the trends observed in Table 1.

The interaction effects are as follows:

1. Dependency has consistent effects upon self-evaluation only when the subject anticipates that his score will be made public. If he has this anticipation, he changes his self-evaluation and ideal score to a greater extent if he

TABLE 1

CHANGE IN SELF-EVALUATION[a]

	Received High Score		Received Low Score	
	Expected High Score	Expected Low Score	Expected High Score	Expected Low Score
Public				
High Dependent	+0.88 (1)[b]	+2.00 (3)	−2.00 (5)	−0.38 (7)
Low Dependent	+0.70 (2)	+1.29 (4)	−0.33 (6)	0.00 (8)
Private				
High Dependent	+0.67 (9)	+0.33 (11)	−0.20 (13)	−0.57 (15)
Low Dependent	+0.40 (10)	+0.75 (12)	−1.50 (14)	−0.75 (16)

[a] The questionnaire item upon which these data are based is:
I think my ability to visualize things in space is

——excellent	——somewhat poor
——very good	——very poor
——good	——extremely poor

The figures indicate the average shift and its direction which occurred as a result of the knowledge of relative standing in the class.
[b] The figures in parentheses are cell numbers to facilitate reference in the text.

TABLE 2

CHANGE IN SCORE THAT WOULD HAVE BEEN OBTAINED UNDER IDEAL CIRCUMSTANCES[a]

	Received High Score		Received Low Score	
	Expected High Score	Expected Low Score	Expected High Score	Expected Low Score
Public				
High Dependent	+8.25 (1)	+25.13 (3)	−28.33 (5)	−21.88 (7)
Low Dependent	+6.40 (2)	+21.14 (4)	−18.44 (6)	−9.25 (8)
Private				
High Dependent	+0.50 (9)	+8.33 (11)	−24.60 (13)	−31.14 (15)
Low Dependent	+0.70 (10)	+6.00 (12)	−33.75 (14)	−30.00 (16)

[a] The sign of the number indicates the direction of the change.

is high than if he is low in his dependence upon others ($p < .01$ by t for both comparisons).[2] There is some indication in Table 1, which is not supported in Table 2, that under the private treatment the reverse effect occurred.

2. Discrepancy produces consistent effects only in the public treatment ($p < .01$ by t for both comparisons). It is only when the subject anticipates confrontation that the larger the discrepancy between expectation and standing the greater the change in both self-evaluation and ideal score. The relationship holds despite an obvious artifact involved in the measure used. The opportunity for change in scale position is in general greater the more the person's performance deviates from his initial expectation. If the subject expects a low score, his self-evaluation should tend to be on the low side. Receiving a very low score further tends to lower his self-evaluation. The opportunity for lowering, however, is generally only one or two scale points at most. On the other hand, if the same subject receives a very high score, the opportunity for increasing his self-evaluation may be as much as five scale points. As a result of the differential ceiling placed upon the possibility for change, therefore, we should expect greater change where expectation and presumed performance are discrepant.

3. When the subject's attained score is high, he is likely to change his self-evaluation (in a positive direction) to a greater extent in the public than in the private treatment. On the other hand, when the subject's attained score is low, he is likely to change his self-evaluation (in the negative direction) to a greater extent in the private treatment ($p < .001$ by t for the interaction). It will be noted that the comparison of Cells 5 with 13 in both tables is an exception to the general trend.

In terms of the distinction between direct and reflected comparison, changes in self-evaluation that occurred in the private treatment may be attributed primarily to direct comparison.[3] An understanding of the effects of anticipated appraisal by others involves interpreting the foregoing interactions.

One of the interactions indicates that dependency showed consistent effects only in the public treatment. Apparently, the degree of dependence upon others for self-evaluation is relevant only where appraisal from others is anticipated. This is the evidence we were looking for in order that we infer the occurrence of reflected comparison in the confrontation treatment.

The interaction between discrepancy and anticipated confrontation suggests that reflected comparison was more important in producing change than was direct comparison. That direct comparison was effective is indicated by the trend, in Table 2, for greater change in the high as compared to the low discrepancy treatments, especially in the high received score treatment. With the addition of anticipated confrontation, however, strong and clear effects emerge.

The interaction between confrontation and received score can best be interpreted by considering the satisfactions mediated by the individual's status in the comparison group. Since the Space Relations test was introduced as a good predictor of academic success, we can assume that the subject considered the skill germane to the activities of the comparison group. The subject's status in the group is presumably mediated by the attribution of skills to him by others in the class. If we assume that

[2] The standard analysis of variance format was not used since many of the comparisons involved collapsing the basic 16-cell design in a number of ways that were not orthogonal. Also, t tests were easier to compute since because of variations in class size (five college classes in all were used) and of attrition, n's were unequal. Comparisons were based upon the amount of change irrespective of direction.

[3] It is possible that in the private treatment some subjects anticipated having to reveal their score at the end of the class session. When a person is asked his score, it is not considered good form for him to refuse to divulge it. To the extent that this occurred, the results for the private treatment are attributable to reflected comparison. It would not be too difficult to work out a procedure in which the operation of reflected comparison would be minimal. A further problem may be suggested at this point. Was the subject's self-evaluation being affected by his relative standing or was change produced as a result of the score discrepancy alone irrespective of where the subject stood in relation to others? This influence could be controlled or eliminated by suitable changes in the experimental procedure. In any future research careful consideration should be given to both of these possible sources of contamination.

high status is desired, and that during the course of the anticipated confrontation high status will be attributed to the subject, he should show little or no resistance to raising his self-evaluation. If, on the other hand, his standing is lower than he anticipated, he should resist lowering his self-evaluation since this would be *de facto* acceptance of lowered status in the group. Since status considerations are not involved in the private treatment, there should be less resistance to lowering self-evaluation. The positive bias observed in the public treatment seems understandable in these terms. Since this experiment was conducted Mischel (1958) has reported a similar finding—that the individual tends to resist lowering his self-evaluation after publicly exposed failure.

Emotionality and Change in Self-Evaluation

A discrepancy between two sources of information about the self (or an aspect thereof) should, by our present assumptions, produce tension which would subsequently tend to be reduced. We have thus far discussed one avenue of tension reduction, namely, a change in self-evaluation that leads to equilibration of the original self-estimate with the estimate derived from social comparison. Presumably, a negative relationship should appear between the amount of change and tension reduction. An indirect test of this consequence was available. The questionnaires administered before and after the Space Relations test contained a question that asked the subject to indicate to what extent emotional feelings tend to interfere on a test like the one he had taken. The amount of change in emotional feelings was found to be negatively correlated with both change in ability evaluation ($p < .05$ by chi square) and change in ideal score ($p < .001$ by chi square).

Other Avenues of Tension Reduction

Various modes of reducing discrepancy are open to the person in the situation studied other than changing his self-evaluation: attributing little validity to the test, denying the importance of the skill in question, attempting subsequently to influence the other to change his appraisal, and shifting the frame of reference for self-evaluation. We shall examine each of these in turn.

Since it is via the test score that the other will come to appraise the individual, devaluing the test as a valid source of information regarding one's ability should also serve to reduce discrepancy tension. Both questionnaires asked the subject to rate the validity of the test as a measure of spatial visualization. Uniformly high validity was attributed to the test on the first questionnaire and very little change in attributed validity occurred, probably because of the way in which the test was introduced and the very elegant and impressive format of the test booklets. There was a slight trend, however, for a subject who expected a high and received a low score to lower his validity estimate more than a subject who both expected and received a low score ($p < .15$ by chi square). If the test were introduced as being of somewhat questionable validity as a measure of the ability, we would expect differences in attributed test validity rather than changes in self-evaluation. To the extent that attributed validity is more labile than self-evaluation we would expect discrepancy tension to produce changes in validity, the results paralleling those obtained in the present experiment.

On both questionnaires the subject was asked to indicate how important the test results were to him. Since the test was described as a good predictor of academic success we can assume that the subject's response reflected the degree to which his score was relevant to skill attribution in his present comparison group. Subjects attributed uniformly high importance to the test. There was some change after the subject discovered his relative standing. When the subject received a low score he tended to lower the importance of the test, whereas when the subject received a high score he tended to increase its importance ($p < .001$ by chi square). Here is evidence for another mode of reducing discrepancy tension.

In discussing change in self-evaluation, the conjecture was advanced that the low scoring subject who anticipated confrontation saw his performance as a threat to his status in the group. His resistance to lowering self-evaluation may reflect the anticipation that he would

be able to influence the other's appraisal of him in subsequent discussion and thus disabuse him of the appraisal reflected by his low score. The prediction should follow of an inverse relation between amount of change and tendency to communicate, an hypothesis inviting further research.

In addition to changes in self-evaluation, changes in scores attributed to members of various groups in the general population were also available for study as an indication of changes in the subject's frame of reference. The subject was asked to guess what score a typical member of the following groups would make: your friends, your class, undergraduates at this school, "the man in the street," and "a first-rate mind." There were no clear effects of dependency and discrepancy. However, the public-private variable produced striking effects that parallel those for self-evaluation and level of aspiration. The attributed scores were consistently lower in the private treatment for all of these groups (in all instances the p values were well beyond the .001 level of confidence). It would appear then that another adjustment to the discrepancy tension induced by anticipated confrontation is a shift in the frame of reference for evaluating one's ability.

SUMMARY

Processes of direct and reflected comparison were distinguished. In the former the person uses the skill evaluation of others as a "yardstick" on which to place himself, whereas in reflected comparison he refers his self-evaluation to the appraisal he assumes others have of him. The study manipulated: the degree of discrepancy between an individual's appraisal of himself and his relative standing in a comparison group, and the anticipation of appraisals from others in the comparison group. The degree to which the person was dependent upon others for self-evaluation was also measured.

Self-appraisal, according to the findings, is influenced by direct comparison of one's performance to the performance of others. The addition of anticipated confrontation (which presumably engenders reflected comparison), produces a positive bias in self-evaluation, produces an increase in the effects of discrepancy, and provides the medium in which differential dependency may operate. Interpretations of these results were offered. Additional avenues of discrepancy tension reduction were also examined.

REFERENCES

COOLEY, C. H. *Social organization.* New York: Scribner, 1909.

DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629–636.

FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92–99.

GERARD, H. B. The effect of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249–271.

MISCHEL, W. The effect of the commitment situation on the generalization of expectancies. *J. Pers.*, 1958, **26**, 508–516.

(Received February 5, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 294-302

THE SOCIAL PSYCHOLOGICAL NATURE OF STRESS: A FRAMEWORK FOR CAUSAL INQUIRY[1]

BRUCE P. DOHRENWEND[2]
Columbia University

A number of recent studies have begun to lay a groundwork of theory and fact about relations between social environmental factors and mental illness. Notable among these researches are the New Haven study (cf. Hollingshead & Redlich, 1958), the Midtown study (cf. Rennie, Srole, Opler, & Langner, 1957), and the Stirling County study (cf. Leighton, 1959). These studies have differed among themselves in several ways: in the definitions of mental disorders employed, in whether they have attempted to assess untreated as well as treated prevalence, in the particular social and cultural factors emphasized, and in the kinds of population studied. The studies have, however, four things in common. First, their emphasis tends on the one hand to be on gross environmental features described, for example, as social disintegration or social class, and on the other hand, on mental disorder. The intervening experiences of the individual receive less theoretical and empirical attention in these searches for the ecological outlines of the problem (cf. Dohrenwend, 1957; Hollingshead & Redlich, 1958, p. 360; Leighton, 1959, pp. 3–5). Second, all set forth as their central assumption, in one form or another, the theoretical position that stressful social and cultural factors are causally related to mental disorder. Third, all report strong relations between various social and cultural factors and mental disorder. And fourth, all are cross-sectional studies of communities or urban neighborhoods, conducted at one time and in one place, without experimental controls.

These last circumstances set severe methodological limitations on the power of these studies to test the general theoretical position. The chief limitation is this: correlations found between sociocultural factors and mental illness in these cross-sectional studies do not demonstrate causal relationships; the well known "drift hypothesis," for example, would suggest an opposite interpretation, counting mental disorder the cause, and disorganized social conditions the effect produced by downward drift of mentally ill persons.

In contrast to these community studies are laboratory investigations of stress. A convergence may be observed between work such as Hans Selye's on physical and chemical changes associated with stress, and such notable research in social psychology as that on frustration and aggression (Dollard, Doob, Miller, Mowrer, & Sears, 1939). Most intriguing in this respect are the recent studies by Funkenstein, King, and Drollette (1957) which attempt to relate psychological stress to physical and chemical changes. Since these studies are controlled experiments, the direction of relationships between the stress situations and their consequences is clear. The difficulty is that the relationships are not between enduring social situations, stress, and mental disorder. The laboratory situations for human subjects must, of necessity, be relatively benign and short-lived. They cannot have, as their aim, the creation of mental disorder. These contrived situations add to a certain artificiality a large degree of impotence compared to stress situations manufactured in social and cultural, environments outside the laboratory. Further the laboratory stress conditions tend to lack even theoretical reference to the very social conditions designated as the major sources of stress in the community studies noted above.

What is needed on an empirical level, then, is sensitivity to stress situations which, though created by social and cultural factors outside the researcher's control, nevertheless provide close approximations to experimental control

[1] This work has been supported in part by Mental Health Project Grant OM-82 from the National Institute of Mental Health, United States Public Health Service. It has been conducted under the joint auspices of the Department of Psychiatry and the School of Public Health and Administrative Medicine.

[2] I am grateful to Barbara S. Dohrenwend, Jack Elinson, and Lawrence C. Kolb for their helpful criticism of this paper.

of meaningful conditions. Such sensitivity, in turn, depends in large part on the existence of theoretical perspective in terms of which these strong and enduring stress situations can be assessed. This paper is an attempt to contribute to such perspective.

THEORETICAL FRAMEWORK

On the basis of their studies of anxiety and stress in paratroop training, Basowitz, Korchin, Persky, and Grinker (1955, pp. 287–289) observe that there are "... difficulties inherent in the prediction of behavior from information about the person and his past experience alone without knowledge of the new setting, its social characteristics, and its meaning to the individual." They conclude that, "In future research ... we should not consider stress as *imposed* upon the organism, but as its response to internal or external processes which reach those threshold levels that strain its physiological and psychological integrative capacities close to or beyond their limits." Thus viewed, what kind of theoretical formulation can be made of the individual's stress response in social and psychological terms?

Clues are available in the work of Hans Selye (cf. 1955, 1956). Selye's investigations have been conducted largely with animals in laboratory settings. His focus has been on physical and chemical changes as manifestations of stress observable in these settings. Nevertheless, elements of his general position provide a useful point of departure for the present endeavor.

Selye's position is that stress is basic to all adaptive reactions. It is thus the common denominator of diverse reactions to a wide variety of agents or "stressors" (the term given by Selye to anything that produces stress). Moreover, the same stressor that has adaptive consequences in one case may have maladaptive results in another. It is this general view of stress that underlies the precise and elegant formulations based in his researches.

It is possible to abstract from Selye's formulations four main elements involved in stress situations: an antecedent stressor (often a toxic agent or electric shock in Selye's work); conditioning or mediating factors such as climate or diet which increase or decrease the impact of the stressor; the General Adaptation Syndrome of nonspecific physical and chemical changes, indicating the intervening state of stress in an organism over time; and consequent adaptive or, in cases where there is a "derailment" of the mechanism of the General Adaptation Syndrome, maladaptive responses, usually in the form of what Selye calls "diseases of adaptation." It is possible to strip these four kinds of elements of particular examples and arrive at the general paradigm of the individual organism's stress response shown in Figure 1 below. This general paradigm, in turn, can be developed in social psychological terms. Consider each of its elements.

Stressor

Since a stressor is any agent that produces stress, the number of stressors is virtually limitless. The concern here is with social stressors—an economic disaster or boom, a clash with the enemy, forced relocation, disruptive contact with representatives of a different culture—these are but a few of the more dramatic and far-reaching social stressors. Reaching fewer people at any given time, but no less important to the individual person, would be the impact of a death, divorce, or

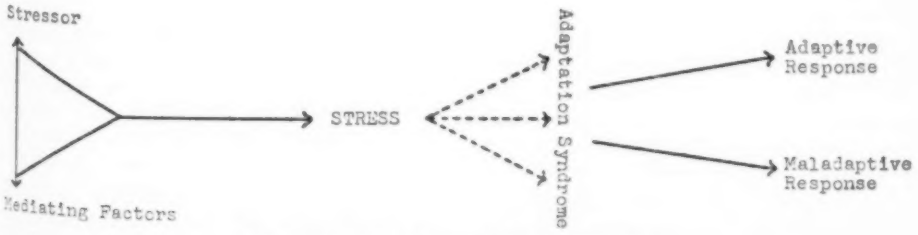

FIG. 1. General paradigm of the stress response.

illness on the members of a family—again, only a few of the possible examples.

Mediating Factors in Relation to a Concept of Constraint

To avoid confusion with the conditioning concepts of learning theory, the term "mediating factors" has been substituted for Selye's "conditioning factors." Selye discusses two different types of such factors: "internal" (i.e., those which have become part of the body through heredity, past experience, etc.) and "external" (i.e., those factors acting from without, such as diet and climate), which influence the response to a physical stressor that is simultaneously acting on the organism. The social and psychological counterparts to be advanced here will be considered under two general headings: first, those mediating factors which determine the amount of *external constraint* associated with the stressor; second, those that determine the amount of *inner constraint* associated with the stressor for the individual in the stress situation.

Constraint is here considered to be the psychological force exerted consciously or unconsciously by the individual to inhibit a course of action called for by events. These events originate in the environment external to the individual or in his own inner drives, desires, and internalized rules or prescriptions. Inner and external constraint are defined in terms of relations between inner and outer events. When force is exerted by the individual to suppress action demanded by inner events in favor of action demanded by outer events, the individual may be said to experience external constraint. When, in contrast, he attempts to inhibit action demanded by outer events in favor of action demanded by inner events, the individual may be said to experience inner constraint. In general, the stronger the external events, the stronger the force to suppress action demanded by internal events in opposition to them, and hence the stronger the external constraint experienced by the individual. Conversely, the stronger the inner events, the stronger the force to suppress action demanded by external events in opposition to them, and hence the stronger the inner constraint experienced by the individual. What, then, are the mediating factors which, in interaction with the stressor, determine the amount and nature of constraint experienced by the individual in the stress situation?

Sources of external constraint. The phenomena of suicide and homicide have been studied as reactions to a most important social stressor, the business cycle. Henry and Short (1954) report some interesting findings on relations between the business cycle and these two extreme actions. These authors show that suicide rates vary inversely and homicide rates vary directly with the degree of what they term "external restraint" associated with this stressor.

Henry and Short describe two major sources of external restraint. One is the status of the individual as based primarily on his ethnic background and occupation; the other is the strength of his "relational system," defined as his social relationships with other individuals, especially his family. In these authors' terms, the lower one's status and the stronger his relational system, the greater the external restraint.

This is a most interesting formulation. It appears, however, to mix two quite intricate things, external *constraint* produced by events related to low status on the one hand, and social *support* associated mainly with a strong relational system (e.g., family ties) on the other. Both are grouped by Henry and Short under the heading "external restraint."

That low status is probably a source of strong external constraint seems apparent. It is, for example, the person who is low in status who is first to be fired in a time of economic decline, thus pressing the previously low ceiling on his aspirations still lower. Whether the individual experiences external constraint or social support for his inner goals from his relational system, however, depends upon a number of factors not considered by Henry and Short.

It seems likely that external constraint increases with the "relative deprivation" experienced by the person as he compares himself with others in the situation (Merton & Kitt, 1950; Stouffer, Suchman, DeVinney, Star, & Williams, 1949); some of these others whose circumstances the individual compares as favorable to his own may be important members of his relational system—his friends,

for example. In addition, to the extent that the person's salient inner goals are dissimilar to those of others in his relational system who are involved in the situation, he is also likely to experience external constraint. It is only when the inner aims of the others are similar to his own that the individual is likely to experience social support as opposed to external constraint from his relational system.

The position adopted for the present formulation, then, is that external constraint varies inversely with status as defined by the wider society, and directly with the relative deprivation experienced by the individual as he compares his position to that of others in the situation. In addition, external constraint varies directly with the dissimilarity between the drives, desires, and internalized prescriptions of the individual and those of the other individuals in the stress situation.

Not only the similarity or dissimilarity of the inner goals of others in such activities are important. The aggregate nature of the rules or norms governing various activities in his relational system should themselves be important as they interpret the stressor to him; for example, if there were a high degree of anomie or normlessness in these activities, one would expect that a confused and ambiguous interpretation of the stressor would be transmitted to the individual by others in his relational system. Thus the quality of external constraint should vary with the state of the norms in the different groups and activities comprising his relational system (Dohrenwend, 1959).

Some examples from a study now in progress may illustrate these points. The subjects of the research are drawn from approximately 1,800 New York City families who found themselves in the path of improvements being made in the approach to a major bridge. These families lived in a narrow strip of eight square blocks bounded by rivers on the west and east. Identical sections of blocks paralleling this relocation section to the north and south provide controls. The stressors of the study, then, consist of various stages of forced relocation of the families in the path of the bridge.

These families come from varying economic and ethnic backgrounds. The western portions of the relocation and control sections have had a plurality of middle class Jewish residents. The central portions, in contrast, tend to be Irish descent and lower middle class. On the east, the plurality is Puerto Rican and economic deterioration is much in evidence. The families to be relocated found themselves face-to-face with the need to move out of rent-controlled apartments into a scarce housing market since no alternative public or private housing was to be built for them. Moreover, the situation within the urban area is such that a move south and east of the relocation section means a move to blocks that are rapidly deteriorating economically. These blocks have also changed in composition from Jewish and Irish descent to Negro and Puerto Rican families in recent years. The threat of ethnic marginality thus faced a Jewish or Irish family considering a move in this direction. One respondent put the situation this way: "The colored and Puerto Ricans are taking over .... My only objection—and I'm not prejudiced—is that the Puerto Ricans refuse to speak English." Another less circumspect informant overstated the change this way: "... once you go down from_____Street and east, your life isn't worth a nickel at 11 o'clock at night."[3]

There are a number of factors in this situation which, as mediators of forced relocation, are probable sources of external constraint for the persons involved. Two factors defining the *social status* of the individual, his class situation and his ethnic background, are of major importance. In a scarce housing market, the individual's opportunities are greatly limited by the rent he can afford to pay. The organization charged with improving the bridge approach set up an office to help relocate the displaced families. By providing bonuses to landlords for renting to such families, this office secured listings of vacant apartments to facilitate relocation. In these circumstances, what an individual is offered depends in large part on his economic resources. The lower his income, the greater the external constraint

[3] These quotations are taken from exploratory interviews with members of families in the relocation section. About 20 respondents were interviewed. I would like to express my appreciation to Claire Jacobson and Warren Swidler who conducted the large majority of these interviews, including those quoted above.

likely to stem from his experiences with the personnel of the Relocation Office and the real estate managers as gatekeepers of the scarce housing market in this situation.

In addition to his class situation, his ethnic background is a key factor. There are discriminatory practices in the real estate market; a favored position is granted to some ethnic groups and an unfavorable position imposed on others. The Relocation Office personnel, as intermediaries between the real estate managers and the families to be relocated, reflected these practices directly by withholding some apartments from some categories of tenants at the behest of the landlord, or indirectly by inability to change the practices of landlords who turned down tenants in these categories. Thus, for example, the darker the prospective tenant's skin or the more pronounced his accent, especially if Spanish, the greater the external constraint he is likely to encounter during relocation.

A person participates, in greater or lesser degree, with other individuals in a wide variety of activities in such spheres as kinship, religion, education, economic exchange, politics, and recreation. In cases where influential persons in one or more of these activities intervened for him with the Relocation Office or with real estate agents, these interventions may have modified—either increased or decreased as the case may be—the external constraint. An active member of a church, for example, might have expected his clergyman to speak on his behalf at the Relocation Office; a small businessman might have had expectations of the local Chamber of Commerce; a tenant might have placed his faith in collective action with other tenants, sponsored perhaps by local political leaders; a wife might have expected her husband to produce a solution; and so on. In general, to the extent that the *inner goals of others* with whom he stands in social relationship are dissimilar to his own, such expectations tend not to be fulfilled; external constraint increases for the individual.

At each stage of relocation, it is possible for the person to compare his own situation with that of others experiencing the same process. At the initial stage when relocation is first announced, for example, it is possible for him to compare what is likely to be available to him with what is likely to be available to his neighbors; later, he is able to compare what the Relocation Office offers his neighbors to what it offers him; he is able to compare the duration of his stay in the relocation section to theirs; and, in some cases at least, he could, still later, compare the quality of their new homes and neighborhoods with his own new location. Further, it is possible for him to compare his economic circumstances prior to relocation with his circumstances after; his old home with his new one; the quality of his old neighborhood with the quality of his new one, in terms of both economic condition and ethnic composition. In short, the person can judge whether his status has increased or decreased; whether his web of social relations has been strengthened or torn asunder. In general, the greater the *relative deprivation* he experiences as he makes these comparisons, the greater the external constraint.

Sources of inner constraint. Quite another, although highly related, set of mediating factors are those that provide the main sources of inner constraint. Of central importance are the person's drives and the internalized rules that give these social direction. These situationally relevant aspects of superego and ego ideal will be referred to hereafter as "self-prescriptions." As they are confronted by a particular stressor, they become sources of inner constraint.

A number of factors should determine the strength of the inner constraint. Major ones would be the sheer number and importance of the individual's self-prescriptions that are affected by the stressor; if, for example, the stressor threatens the self-prescriptions that define the person's status aspirations as well as other aspects of his self-image, he is likely to experience greater inner constraint than if such aspirations are not in jeopardy. The number and importance of others with similar self-prescriptions should also play a large part; if other such persons are in the same situation, interaction with them should strengthen his self-prescriptions and hence the inner constraint he experiences. In this view, social support functions to increase inner constraint. Moreover, to the extent that a person's self-prescriptions are consistent and support each other, the aggregate inner constraint is likely to be increased.

And, of great importance, the stronger per se the self-prescriptions governing the expression of the person's drives, the stronger the inner constraint he is likely to experience. Thus, in general, the strength of self-prescriptions and hence inner constraint may be greater in the case of the individual with the stronger superego stemming from childhood discipline by withheld love and approval rather than physical punishment (Sears, Maccoby, & Levin, 1957). There is evidence that these differing types of childhood discipline vary with social class (cf. Bronfenbrenner, 1958). They may also vary with ethnicity.

Consider again the person faced with forced relocation. Should relocation involve his status aspirations, his business interests, his hopes for the education of his children, his internalized obligations to relatives and friend—a large *number of his self-prescriptions* as opposed to very few—the inner constraint is likely to be greater. Thus, for example, a person who has business interests in the area is likely to experience greater inner constraint than one who does not, other things constant.

In addition, the *number and importance of others with similar self-prescriptions* with whom the person interacts should affect his experienced inner constraint. The assumption is that the more such persons there are, and the more important they are to the individual, the more likely interaction with them strengthens his internalized aims and obligations. Thus, for example, if his spouse had the same neighborhood groups as points of positive and/or negative reference, his self-prescriptions with regard to these groups are likely to be strengthened by his interaction with his spouse.

It is likely that *socialization practices* vary with both class and ethnic background. Some of these variations may affect the strength of the rules internalized by individuals—their self-prescriptions—and hence the inner constraint experienced. Thus, for example, should it be found that physical punishment is reported as a more frequent socialization practice in lower than in middle class families of a given ethnic background, inner constraint may also be weaker for persons of the former rather than the latter background, other things constant.

A note on other mediating factors. Other factors than those just considered undoubtedly also mediate the impact of the stressor on the individual. Differences in intellectual abilities and physical health, for example, have to be considered as factors affecting the constraint experienced. So also do differences in age and sex. The emphasis above has been on factors with which the theory has been able to deal somewhat systematically at its present stage of development. Their formulation is not meant to indicate closure that would be, at this point, premature.

Adaptation Syndrome

The stressor, in association with mediating factors, then, produces inner and external constraint. The person's subsequent attempts to adapt to constraint are represented by what is termed his adaptation syndrome. A central assumption of the theory is that the adaptation syndrome represents efforts by the individual to reduce constraint.

A question arises as to the manner and degree to which something like the defense mechanisms postulated in psychoanalytic theory (cf. Freud, 1946) underlie adaptation syndromes. For the time being, this relevant and important question will be left open. For the present, three broad, conceptually distinct facets of social psychological adaptation to constraint may be considered: the affective, the conative, and the cognitive. Each will be developed in turn.

Affective facets. Basic affective aspects of adaptation to constraint are fear and anger, and that complex combination of the two, anxiety. For human organisms in social situations, these can become highly elaborated. Of special interest in this regard is the "frustration-aggression hypothesis" investigated at Yale (Dollard et al., 1939), and more recent studies such as those of Henry and Short (1954) and Funkenstein et al. (1957). The latter two studies have centered on the direction taken by anger consequent upon frustrating circumstances—"anger-in," that is, anger directed by the person towards himself; and "anger-out," or anger directed towards objects, especially other individuals, in his environment. This emphasis on the direction of anger takes on additional significance in light of some of the research described by Funkenstein and his associates (1957), who suggest that transient

physical and chemical correlates of anger-in for healthy subjects resemble conditions found to be chronic in depressive psychotics. In contrast, the correlates of anger-out in healthy subjects tend to be similar to chronic physical and chemical states in paranoid schizophrenics.

Conative facets. Adaptation to constraint involves changes in the person's social activity. His participation in any given social interaction in which he has been engaged may increase or decrease; some activities may be abandoned; new activities may be undertaken. Old or new activities may be undertaken with other persons the individual has known in the past, or with others with whom he has had no prior acquaintance. The activity may be governed primarily by affect or by social rules; the latter may be either his own internalized rules, or rules of others that are incongruent with them.

Cognitive facets. Changes should also occur in the orientation of the individual as he adapts to constraint. Certain of his self-prescriptions, for example, may take on heightened importance and strength; in short, they may be reaffirmed. In contrast, the self-prescriptions may become less important, weaker—in which case he may be described as becoming alienated from them. Or, there may be a change in these self-prescriptions amounting to reorientation (Dohrenwend & Smith, 1957).

The problem of baseline combinations of these affective, conative, and cognitive facets of adaptation. There is no ready answer to the question of how these various affective, conative, and cognitive facets of adaptation combine to form one or more baseline syndromes of social psychological adaptation, syndromes that serve as relatively neutral standards against which adaptive vs. maladaptive departures can be assessed. The nature of the problem can be illustrated by considering clinical descriptions of reactions to combat, perhaps the best documented of stress situations in psychological terms.

"Normal battle reaction" is described by Noyes and Kolb (1958, pp. 532–534) in the following way: There is complete and continuous alertness, an alertness associated with fear, even terror, and often with anger in the form of irritability and resentment; these can bring the person to the point of exhaustion; commonly accompanying these reactions are a number of psychosomatic disturbances—excessive perspiration, nausea, difficulty sleeping, muscular weakness, oversensitivity to threatening noises are a few examples.

Noyes and Kolb (1958) suggest that these reactions can be considered a baseline against which pathological (maladaptive) responses should be evaluated. This baseline is, however, a relative one. Such baseline reactions, these authors state, " . . . would be considered abnormal in a civilian setting yet are not so incapacitating as to demand removal from combat. Unless extreme, such reactions must *under the circumstances* be considered within the range of normal" (p. 533, italics added). Implicit is the assumption that the reactions have not been combined with changes of orientation or social interaction that would seriously undermine "morale," defined as " . . . those attitudes, feelings and beliefs which promote participation in a united effort" (p. 529).

The relative nature of the "normal battle reaction" as a baseline for assessing whether departures are maladaptive underlines a more general question, the question of whether it will be possible to speak of a *general* adaptation syndrome in meaningful social psychological terms. In large part, of course, this is a problem for research. For the present, however, it seems reasonable to take the position that adaptation may have a number of baseline social psychological manifestations which form syndromes that are distinct and different from each other both for different persons in the same general situation, and for the same person in different situations.

Adaptive and Maladaptive Responses

A number of extreme responses to combat have come to be taken as indications of "combat neurosis" (Noyes & Kolb, 1958, pp. 533–535). One form of this neurosis is termed an anxiety state; in this state, fear becomes panic which is manifested in flight, "freezing," or "primitive protective withdrawal, with stupor, catatonia, and mutism." Another form of combat neurosis consists in conversion reactions, with, for example, sensory losses, aphonia, deafness, partial or total blindness and occasionally, amnesia. Still a third form of this neurosis is depression characterized by

extreme anger-in, evidenced by ideas of failure, guilt and self-condemnation behind a façade of "... rigid, mask-like face, staring eyes that wink infrequently, ... paucity of movement ...." These extremes of fear, anger-in, weakening of self-prescriptions, and withdrawal from social activities, necessitate removal of the individual from combat. They are hence maladaptive within a military frame of reference, and would be considered maladaptive in most other contexts as well. In short, there is strong indication that the mechanisms underlying adaptation have become, in Selye's term, "derailed," leading to maladaptive stress responses.

In contrast are the adaptive responses to combat. Within a military frame of reference, anger-out, if directed at the enemy, is considered an adaptive response; so also is reaffirmation of collectivistically oriented self-prescriptions which dictate increased loyalty, obligation to other soldiers, and greater participation in the group effort. Such responses add up to increased military effectiveness and heightened morale. Yet such responses would by no means be considered adaptive in all other circumstances—situations, for example, in which killing as many members of the outgroup as possible is not a social goal of the action. And the tremendous range of responses between the extremes of incapacitating combat neurosis as against heightened morale further compound the dilemma of ascertaining the grounds on which, and the degree to which, a stress response is adaptive or maladaptive.

Hollingshead and Redlich (1958, p. 11, 179) state of their New Haven research that, in dealing only with psychiatrically treated mental illness, they are assuming a social definition of mental disorder. For the present, it may make most sense to broaden and extend this idea of a social definition of adaptive and maladaptive behavior by widening the range of sources of such definitions. Thus, for example, several agencies of law, health, and welfare in our society have, as their primary purpose, defining whether or not the responses of individuals who come to their attention are adaptive or maladaptive over a wide range of circumstances. Where a person is of school age, educational agencies bring their perspective to bear. In addition, a person's primary groups pass judgment on whether his behavior is adaptive or maladaptive. Finally, the individual himself, implicitly or explicitly, takes stock of his own behavior. In each case, the frame of reference used for such assessments consists of the rules or standards by which the agency, group, or individual operates.

Note that, in this formulation, it is by no means necessary for the various definitions of adaptive vs. maladaptive responses to agree. A person, for example, does not have to consider himself as suffering in order to be diagnosed as mentally ill. Moreover, different responses by the same individual may be defined as adaptive in one case, and maladaptive in another, by the same source of assessment. Thus a theft may be thought adaptive by the thief's primary group, who judge as maladaptive only his carelessness at being caught. In this approach, the focus is on mental disorder within a wider framework of social deviance.

SOME PROPOSITIONS

The theory as thus far developed can be summarized by a number of assumptions about the general nature of relationships among the various elements of the stress response. The assumptions refer, of course, to classes of stressors that are primarily social in nature and do not involve direct and immediate physical harm to the individual. Stated in propositional form, they are as follows:

Proposition 1. Stress is a state intervening between antecedent constraint and consequent efforts to reduce constraint.

Proposition 2. Stress underlies both adaptive and maladaptive behavior.

Proposition 3. The probability of maladaptive responses varies directly with the intensity *and* duration of stress.

Proposition 4. The intensity of stress varies directly with the strength of external constraint associated with the stressor.

Proposition 5. The intensity of stress varies directly with the strength of inner constraint associated with the stressor.

Proposition 6. The adaptation syndrome represents efforts to reduce constraint.

Proposition 7. The nature of the adaptation

syndrome varies with: the *relative* strengths of external constraint vs. inner constraint associated with the stressor, and the nature of the sources of inner and external constraint.

Proposition 8. The nature of the adaptive and/or maladaptive responses varies with the nature of the adaptation syndrome.

REFERENCES

BASOWITZ, H., KORCHIN, S. J., PERSKY, H., & GRINKER, R. R. *Anxiety and stress.* New York: McGraw-Hill, 1955.

BRONFENBRENNER, U. Socialization and social class through time and space. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, 1958. Pp. 400–425.

DOHRENWEND, B. P. The Stirling County study. *Amer. Psychologist,* 1957, **12,** 78–85.

DOHRENWEND, B. P. Egoism, altruism, anomie and fatalism: A conceptual analysis of Durkheim's types. *Amer. sociol. Rev.,* 1959, **24,** 466–473.

DOHRENWEND, B. P., & SMITH, R. J. A suggested framework for the study of acculturation. In V. F. Ray (Ed.), *Cultural stability and cultural change.* Seattle: American Ethnological Society, 1957. Pp. 76–83.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. F. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

FREUD, Anna. *The ego and the mechanisms of defense.* New York: International Univer. Press, 1946.

FUNKENSTEIN, D. F., KING, S. H., & DROLETTE, MARGARET E. *Mastery of stress.* Cambridge: Harvard Univer. Press, 1957.

HENRY, A. F., & SHORT, J. F. *Suicide and homicide.* Glencoe, Ill.: Free Press, 1954.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness.* New York: Wiley, 1958.

LEIGHTON, A. H. *My name is legion.* New York: Basic Books, 1959.

MERTON, R. K., & KITT, ALICE S. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Continuities in social research.* Glencoe, Ill.: Free Press, 1950. Pp. 40–105.

NOYES, A. P., & KOLB, L. C. *Modern clinical psychiatry.* (5th ed.) Philadelphia & London: Saunders, 1958.

RENNIE, T. A. C., SROLE, L., OPLER, M. K., & LANGNER, T. S. Urban life and mental health. *Amer. J. Psychiat.,* 1957, **113,** 831–836.

SEARS, R. R., MACCOBY, ELEANOR, E., & LEVIN, H. *Patterns of child rearing.* Evanston, Ill.: Row, Peterson, 1957.

SELYE, H. Stress and disease. *Science,* 1955, **122,** 625–631.

SELYE, H. *The stress of life.* New York: McGraw-Hill, 1956.

STOUFFER, S. A., SUCHMAN, E. A., DEVINNEY, L. C., STAR, SHIRLEY A., & WILLIAMS, R. M., JR. *The American soldier.* Vol. I. Princeton, N. J.: Princeton Univer. Press, 1949.

(Received February 11, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 303-308

INDIVIDUAL DIFFERENCES IN REACTION TO FAILURE: MODE OF COPING WITH ANXIETY AND INTERFERENCE PRONENESS[1]

ALFRED LOWE[2]
Boston University

This is an attempt to investigate and clarify some of the sources of individual differences in reaction to stress as manifested in ongoing performance on a cognitive task. This problem is approached through a conceptualization of stress (or frustration) suggested by Lazarus, Deese, and Osler (1952), Sargent (1948), and others, which places the arousal of affect or anxiety at the center of the psychological experience resulting from interference with the pursuit of a goal. The behavioral consequences in terms of ongoing performance are understood to be determined to a large degree by the effects of the individual's efforts to cope with the aroused affect, and by his ability to withstand and overcome its interference in cognitive functioning (Lowe, 1959).

Two contrasting modes of coping with anxiety aroused by frustration were selected for study: the "goal oriented" mode, in which continued pursuit of the blocked goal is the path chosen for the reduction of affect; and the "ego oriented" mode, in which withdrawal from the stressful situation is instrumental towards the reduction of tension. In terms of behavior, the former mode may be expected to lead to an increase of efforts invested in the task and the latter to a relative decline.

These reactions are based on concepts described by Rosenzweig (1943) and Bloom (1952). While it is understood that there are other modes of tension reduction and that people are able to use more than one mode, it is assumed (and there is much support in the literature for this assumption) that individuals develop preferred modes of coping with anxiety and that these become relatively stable personality characteristics.

In addition to this indirect effect on behavior via the mode of coping, anxiety also affects ongoing performance directly, as an interfering, distracting stimulus. The ability to resist and overcome such interference varies between individuals, and has been conceptualized here as interference proneness.

This study then deals with two independent variables: mode of coping with anxiety, and interference proneness. Each has its main impact on a specific facet of cognitive functioning: the former operates essentially through shifts in motivation, and therefore its behavioral manifestation must be sought in the consequences of differential energy output. The latter, as other kinds of distraction, may be expected to affect mainly the accuracy of performance.

The general proposition can be stated as follows: The effect of stress on ongoing performance is to an important degree determined by the individual's mode of coping with the anxiety aroused and by his ability to resist or overcome its interference in cognitive functioning.

The following predictions were derived from this statement and the preceding theoretical considerations: (*a*) Subjects using the goal oriented mode of coping with anxiety show larger output on an output task under stress than prior to the induction of stress, whereas subjects using the ego oriented mode of coping with anxiety show no such increase in output. (*b*) Subjects classified as highly interference prone make more errors on a cognitive task under stress than prior to its induction, whereas subjects classified as low in interference proneness do not show this tendency.

METHOD

The appropriate selection of procedures for the induction of stress, of measures for the variables dealt with, and of experimental tasks relevant to these

[1] This article is based on a dissertation submitted to the Graduate School, Boston University, in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

[2] The author wishes to express his appreciation to Austin W. Berkeley, A. W. Hire, and W. Weiss for their assistance and advice, and to John Arsenian for his aid and encouragement.

At time of writing with the Boston State Hospital.

variables constitutes a major problem confronting this kind of research. In the present study, measures were required to assess the subjects' mode of coping with anxiety and their proneness to the interference of anxiety in cognitive functioning. The experimental task had to be sensitive to the effects of a change in motivation as well as of distraction.

Induction of Stress

Stress was induced by exposing the subjects to failure on a task on which they were motivated to do well. This task was represented as an intelligence test, which led to considerable involvement by the subjects, with many signs of anxiety and concern. In view of this and the fact that similar methods have often been used in related studies, the assumption that the subjects were actually frustrated by failure seems warranted.

Mode of Coping with Anxiety

In order to differentiate the subjects into goal oriented and ego oriented categories, a measure was needed that would reveal differences in motivation under frustration. Such an instrument is provided by the selective recall phenomenon. Previous studies have shown that the direction of differential recall, i.e., the relative preponderance of successes or failures in recall, are related to motivational characteristics. Bloom (1952) confirmed the operation of personality factors in differential recall by demonstrating the existence of a relationship between differential memory for failures and successes on the one hand and changes in motivation on the other. Subjects who recalled a preponderance of failures tended to increase their motivation on a level of aspiration experiment following failure, while subjects recalling a preponderance of successes lowered their goals following failure. A questionnaire further confirmed this relationship. Bloom's study amounts to a validation of differential recall as a measure of personality characteristics which are closely related to modes of coping with anxiety as conceptualized here. Other studies reporting results consistent with the above are those of Atkinson (1953) and McClelland and Liberman (1949).

Accordingly, the subjects' preferred mode of coping with anxiety was inferred from their relative emphasis in recall on successes and failures. In order to control for possible significance of the size of total recall, the actual score used was the ratio of the number of failures recalled to total recall. This will be referred to as the failure recall ratio (FRR). Thus subjects above the median, or upper quartile, on this measure were operationally defined as goal oriented, and subjects below the median, or lower quartile, on this ratio were operationally defined as ego oriented.

The experimental task which was used for the induction of failure was simultaneously used for the measurement of differential recall preferences. It consisted of 14 scrambled sentences: 7 could be meaningfully rearranged, but the other 7 were insoluble. The sentences used in this experiment were taken from two lists of 14 sentences each used by Caron (1954) for a similar purpose. They consisted of approximately 20 words each bunched in phrases of 2 or 3 words; these phrases were scrambled and had to be rearranged by the subject so as to make sense. In order to facilitate recall, each sentence was given a name by selecting the phrase that best represented the gist of the sentence and placing it, underlined, at its beginning.

This kind of task is well suited for the purpose at hand. It has a certain interest for the subject, is credible as a measure of intelligence, and lends itself easily to the manipulations it is subjected to. Levels of difficulty can be adjusted to the abilities of the subjects and recall can be easily ascertained.

Interference Proneness

A scale composed of 19 items culled from a questionnaire used by Waterhouse and Child (1953) in a study dealing with interference proneness was used for the assessment of the subjects' susceptibility to the interference of anxiety in cognitive functioning. In order to maximize the concept validity of the scale, the items relevant to this notion were selected independently by four judges on the basis of the assumption subjects with a score above the median would be considered high on interference proneness and subjects with scores below the median would be considered low. This measure is referred to as the interference score (IS).

Performance Measures

The task to be set had to be sensitive to shifts in motivation as well as to the effects of interference. Study of the former generally involves operations that permit inferences as to the amount of energy expended by the subject, and a common method is the use of simple output tasks (Atkinson & Raphelson, 1956; Lowell, 1952).

With regard to the second requirement, that the task be sensitive to the effects of interference, the literature provides less clear guidance. It seems that the effects of interference from inner sources on ongoing performance have never been thoroughly investigated. Rapaport (Rapaport, Gill, & Schafer, 1945) touches on this problem when he considers that "attention" is hampered "if the subject's affect and anxieties are not well controlled and get out of balance." It certainly seems likely that there are special aspects of behavior which are particularly vulnerable to the operation of interfering stimuli. While these aspects undoubtedly are dependent on the nature and complexity of the task as well as on the intensity of the interference, it is likely that one important effect of distraction will be a decline in the accuracy of execution of the task, i.e., an increase in errors.

In view of the above considerations, it was decided to employ output or speed of performance as a measure of changes in motivation, and errors as a measure of interference. A suitable instrument that combined both measures was at hand in the Minnesota Clerical Test (MCT) (Andrew & Paterson, 1946). This is a test of speed and accuracy which involves the scanning of pairs of numbers and of names and the detection of minor differences between the members of a pair. In half the pairs the members are identical; in the other half, distributed at random, the members of the pairs are different. Two scores were compiled for each subject: an output score, consisting of the number of items

attempted, and an error score, consisting of the items on which errors were made.[3]

Subjects and Procedure

Sixty-five student nurses, required to participate, served as subjects. In order to rule out intelligence as a relevant variable, the relationship between differential recall and level of intellectual functioning was explored. It was found to be not significant.

Each subject was seen twice; first in a group setting, when the questionnaire was administered, and then individually for approximately one hour divided in the following way:

1. Rearranging the seven soluble scrambled sentences, "A"
2. The first half of each part of the MCT
3. Recall of sentences "A"
4. Rearranging the seven soluble scrambled sentences, "B"
5. The second half of each part of the MCT
6. Recall of sentences "B"

The introduction to the research as well as the instructions to the various tests were phrased so as to induce high motivation in the subjects to do well. In order to get the subjects ego involved for the subjective experience of failure to come about, the tests were represented as intelligence tests.

The following are the null hypotheses tested:

1. The mean difference in output postfailure minus prefailure for the group of subjects above the median on the FRR is equal to that of the subjects below the median on this measure, i.e., there is no interaction between experimental conditions and recall. The same relationship holds for the extreme groups defined by the upper and lower quartiles.

2. The mean difference in errors postfailure minus prefailure for the group of subjects above the median on the IS is equal to that of the subjects below the median on this measure, i.e., there is no interaction between experimental conditions and interference proneness. The same relationship holds for the extreme groups defined by the upper and lower quartiles.

The predictions mentioned earlier are the alternatives to these null hypotheses.

RESULTS

For the purposes of analysis, subjects were divided into high and low groups on the basis of their FRR, and these groups were further subdivided into high and low groups on the basis of their IS. Four experimental groups were thus differentiated:

1. High FRR–High IS
2. High FRR–Low IS
3. Low FRR–High IS
4. Low FRR–Low IS

To test the different hypotheses, the high and low groups were alternately defined by the median of the respective distributions and by the upper and lower quartiles.

To test the significance of the relationships obtained, the output scores and the error scores were subjected to a complete analysis of variance in a three-way factorial design which took into account the correlation arising out of repeated measurement of the subjects before and after failure.

The results of the experiment generally, with one exception, support the predictions made and permit rejection of the null hypotheses.

Table 1 shows that the subjects high on FRR have significantly larger output postfailure, while there is no evidence of difference in the output of subjects low on FRR. The difference between the postminus-prefailure output differentials of the two groups is highly significant. That for the high FRR group is more than twice that of the low FRR group. These relations obtain both when the median and when the extreme quartiles are taken to define the groups.

Table 2 summarizes the analysis of variance of the mean output scores pre- and postfailure of the four experimental groups when defined by the medians of the distributions. The fifth row of this table shows the significance of the variance contributed by the interaction of experimental conditions with recall characteristics to the total variance of output scores.

TABLE 1

MEAN OUTPUT SCORES OF SUBJECTS ABOVE AND BELOW MEDIAN FRR VALUE PRE- AND POSTFAILURE, AND MEAN DIFFERENCES BETWEEN EXPERIMENTAL CONDITIONS

Failure Recall Ratio	Prefailure	Postfailure	(Post-Pre) Failure
Above Median (n = 29)	139.44	150.55	11.11*
Below Median (n = 36)	128.30	133.05	4.75**
("Above" minus "Below")			6.36***

Note.—Probabilities are based on one-tailed tests of significance.

* p = .05.
** p = .02.
*** p = .002.

[3] Correlations among the principal measures were not significant. For speed and accuracy in the MCT, the r was .11. For the indices of interference proneness and mode of coping with stress, the correlation was .15. In the absence of evidence that these scores entail any serious overlap, it seems quite legitimate to assume that the processes as defined are relatively independent.

TABLE 2

ANALYSIS OF VARIANCE OF OUTPUT SCORES OBTAINED PRE- AND POSTFAILURE BY FOUR GROUPS OF SUBJECTS CLASSIFIED BY THE MEDIAN FRR AND IS VALUES

Source	*SS*	*df*	*F*
Recall	12.024	1	6.60[a]
Interference	3.429	1	1.88
Condition	3.243	1	58.11[b]
R × I	0.948	1	
R × C	0.577	1	10.34[b]
I × C	0.156	1	2.79
R × I × C	0.030	1	
Total Among	20.347	7	
Subject	111.062	61	
S × C	3.404	61	
Total Within	114.466	122	
Total	134.813	129	

Note.—The analysis is based on the square roots of the raw scores.

[a] $F.95 = 4.40$.

[b] $F.99 = 7.08$.

The same probability of such interaction occurring, less than .01, attaches to the extreme groups.

With regard to the interference hypothesis, no significant differences were found between experimental conditions or between groups when they were defined by the median scores on the IS. As Table 3 shows, when the analysis of the data was confined to the extreme groups (those above the upper and below the lower quartiles on the IS), it was found that the high interference prone subjects significantly increased the number of errors after failure as compared to their performance before failure, whereas the low interference prone subjects actually reduced the number of errors from prefailure to postfailure.

Table 4 presents the summary of the analysis of variance of the error scores obtained pre- and postfailure by the extreme IS groups subdivided by the subjects' standing with regard to the median of the FRR. The fifth

TABLE 3

MEAN ERROR SCORES OF SUBJECTS ABOVE THE UPPER AND BELOW THE LOWER IS QUARTILES PRE- AND POSTFAILURE, AND THE MEAN DIFFERENCES BETWEEN EXPERIMENTAL CONDITIONS

IS	Pre-failure	Post-failure	(Post-Pre) Failure
Upper Quartile (n = 16)	4.81	6.12	1.31
Lower Quartile	4.25	2.81	−1.44
("Upper" minus "Lower")			2.75

TABLE 4

ANALYSIS OF VARIANCE OF ERROR SCORES OBTAINED PRE- AND POSTFAILURE BY THE EXTREME IS GROUPS (Upper and Lower Quartiles), SUBDIVIDED AT THE MEDIAN FRR VALUE

Source	*SS*	*df*	*F*
Interference	3.06	1	3.55[a]
Recall	5.10	1	5.93[a]
Condition	0.22	1	
I × R	2.26	1	2.62
I × C	1.54	1	6.41[a]
R × C	0.46	1	1.91
I × R × C	0.04	1	
Total Among	12.60	7	
Subject	24.15	28	
S × C	6.74	28	
Total Within	30.89	56	
Total	43.49	63	

Note.—The analysis is based on the square roots of the raw scores plus .5 ($\sqrt{x} + .5$).

[a] $F.95 = 4.2$.

row gives the variance due to the interaction of experimental conditions with the interference variable. The probability of such an interaction occurring is less than .02.

The null hypotheses stipulating no interaction between (*a*) experimental conditions and modes of coping with anxiety and (*b*) experimental conditions and interference proneness could therefore be rejected with confidence.

DISCUSSION

The general proposition advanced earlier finds much support in the results obtained. Subjects classified as goal oriented in coping with anxiety increased output significantly following failure, whereas the subjects categorized as ego oriented showed no significant increase. Similarly the group classified as highly prone to the interference of anxiety in cognitive functioning registered an increase in the number of errors after failure, while the least interference prone group actually showed a decrease.

The finding that the predicted relationships emerged, on the whole—particularly where the interference variable was concerned—more clearly when the extreme groups alone were considered, suggests that other, so far obscure, factors enter into the situation, which the tendencies postulated here are able to overcome only when they are at their strongest. Subjects in the middle ranges may be more flexible in their reaction to failure. This would be consistent with Bloom's (1952) finding that subjects who had a balanced recall pat-

tern differed from the extreme groups in (among other ways) a more flexible approach to the frustrating situations facing them. Flexibility implies either the ability to use different coping mechanisms without marked preference for any one or inconsistency in their use. Such a trait could thus account for the greater variability of the scores obtained from the subjects in the middle ranges on the FRR.

The fact that the intermediate ranges of the IS fail to differentiate groups meaningfully may be due to the intrusion of other factors not considered here, such as the intensity of stress. It is probable that more severe anxiety would evoke different reactions in the predicted directions over the entire range.

The dual role of anxiety in its effect on performance under stress emerges clearly in this study. In its role as signal, it determines the fate of motivation and thus indirectly of performance. In this capacity it affects primarily the quantitative aspects of the task. In its other role as the active interfering agent, it has a direct effect on performance and impinges mainly on its qualitative aspects.

The results emphasize the necessity to distinguish between the different effects of anxiety in an evaluation of the consequences of frustration on performance. It is apparent from the differential vulnerability of certain aspects of functioning and from the immunity of others that these effects are specific in their consequences and relatively independent from one another. Many conflicting and inconclusive findings in this general area may be due to a failure to take this distinction between direct and indirect effects of anxiety and their different relevance to specific areas of functioning into consideration.

As in many other studies mainly concerned with discovering changes in performance arising out of the introduction of stress per se, a main effect, ascribable to experimental conditions, was found in addition to the interaction effect discussed earlier. This consists of an increase of output following failure, which, as Table 2 shows, is statistically highly significant. No such consequences of failure were noted on the number of errors.

While the majority of subjects thus increased output following failure, the relative extent of this increase is, as the results show, firmly related to differences in the postulated defensive processes. Also, a substantial proportion of subjects (over 18%) did not increase output at all under failure, and—what is more significant—of these, over 83% were classified as ego oriented on the basis of their FRR. Conversely, only 7% of the goal oriented group failed to increase output following failure. Although this is in accord with theoretical expectations, it still seems profitable to speculate as to the probable reasons for the actual increase of output after failure of a large part of the group resorting to the ego oriented mode of coping with anxiety.

The relative mildness of the stress almost certainly is a relevant factor, and it is likely that with more severe anxiety, the tendency towards withdrawal, thought to be inherent in this mode of coping, would be more accentuated. Other elements in the total situation also act against leaving the field on the part of the subject and thus obscure the consequences of the defensive reaction. Prominent among these are the effects of the instructions and the wish to please the experimenter. This extrinsic motivation can be thought of as to some extent counteracting the ego oriented subject's inclination towards withdrawal from the situation, so that investment in the task is maintained on a level high enough to keep the subject from reducing his efforts and possibly lead to some increase in output. Perhaps we are confronted here with some of the processes subsumed under the concept "competence" by White (1959). Under more severe stress, some of these factors are likely to lose their force, and relatively neutral compliance can be expected to turn into withdrawal. This is a question that should find an answer in more experimentation.

In its conceptualization, operations, and results, this study represents an integration of a number of phenomena which have been the objects of numerous unrelated investigations. It forms a bridge between the large number of studies of changes in performance in response to stress, on the one hand, and of selective recall of successes and failures on the other. By demonstrating that a relation exists between changes in performance following failure and differential recall of successes and failures, it is possible to impart a new and wider significance to this phenomenon. In an

indirect way, through Bloom's (1952) finding that selective recall is related to behavior in level of aspiration experiments, the present study is also linked to that problem area. Since performance following failure has been shown to be related to selective recall and the latter to level of aspiration, changes in performance under stress should be positively correlated with shifts in level of aspiration following failure. In other words, common personality factors can be assumed to underlie these reactions, which were here conceptualized as modes of coping with anxiety.

Of theoretical as well as methodological significance is the demonstration that frustration has not a unitary effect on ongoing performance which can be assessed by means of some composite measure of improvement or deterioration. Rather, different facets of behavior seem to be affected differently according to individual variations in the constellation of personality characteristics. Improvement or deterioration under stress is not necessarily uniform, and in assessing the consequences of the introduction of stress, it is necessary to specify which aspects of performance are improved or impaired. This study has been able to discriminate between varying effects of failure on both quantitative aspects of ongoing performance (output) and qualitative aspects (errors) and to specify some of the sources of these differences in reaction.

SUMMARY

The study begins with the proposition that the individual's mode of coping with anxiety and his ability to resist and overcome the interference of anxiety in cognitive functioning are important determinants of performance under stress. Subjects were divided into goal oriented and ego oriented groups and subdivided into high and low interference prone groups. Various aspects of performance on a cognitive task preceding and following failure in a situation in which they were motivated to succeed were compared.

The results supported the hypotheses (*a*) that the goal oriented subjects increase output significantly following failure, whereas the ego oriented subjects do not, and (*b*) that the high interference prone subjects make more errors after failure than before, whereas the low interference prone subjects do not.

Some of the implications of these results for the general problem of individual differences in performance under stress were discussed.

REFERENCES

ANDREW, D. M., & PATERSON, D. G. *Minnesota Clerical Test: Manual*. New York: Psychological Corporation, 1946.

ATKINSON, J. W. The achievement motive and recall of interrupted and completed tasks. *J. exp. Psychol.*, 1953, **46**, 381–390.

ATKINSON, J. W., & RAPHELSON, A. C. Individual differences in motivation and behavior in particular situations. *J. Pers.*, 1956, **24**, 349–363.

BLOOM, B. L. Personality structure and the recall of interrupted tasks. Unpublished doctoral dissertation, University of Connecticut, 1952.

CARON, A. J. Effects of task, objective orientation and personality factors on recall in an interrupted task situation. Unpublished doctoral dissertation, University of Chicago, 1954.

LAZARUS, R. S., DEESE, J., & OSLER, SONIA F. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293–317.

LOWE, A. Individual differences in reaction to failure: Mode of coping with anxiety, and interference-proneness. Unpublished doctoral dissertation, Boston University, 1959.

LOWELL, E. L. The effect of need for achievement on learning and speed of performance. *J. Psychol.*, 1952, **33**, 31–40.

MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement in recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.

RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. I. Chicago: Year Book Publishers, 1945.

ROSENZWEIG, S. An experimental study of 'repression' with special reference to need-persistive and ego-defensive reactions to frustration. *J. exp. Psychol.*, 1943, **32**, 64–74.

SARGENT, S. S. Reaction to frustration—A critique and hypothesis. *Psychol. Rev.*, 1948, **55**, 108–114.

WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: II. An experimental study. *J. Pers.*, 1953, **21**, 298–311.

WHITE, R. W. Motivation reconsidered: The concept of competence. *Psychol. Rev.*, 1959, **66**, 297–333.

(Received February 11, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 309-314

SELF-CONCEPT, MOTHER CONCEPT, AND FOOD AVERSIONS IN EMOTIONALLY DISTURBED AND NORMAL CHILDREN

ANTHONY DAVIDS[1] AND MARCIA J. LAWTON[2]
Brown University and Emma Pendleton Bradley Hospital

In recent years an increasing amount of research has been devoted to studies of mother-child relations. Psychoanalytic theory and most current theories of child development have stressed the crucial role played by early experiences in influencing cognitive processes and personality formation. The focus of the present study is on the following three important phenomena that are believed to be influenced by mother-child relations: the child's self-concept, the child's concept of the mother, and the child's food aversions. The purpose of the investigation is to study interrelations among these three factors in a group of normal children and in a group of hospitalized emotionally disturbed children, and to compare the two groups on each of these variables.

More specifically, on the basis of psychoanalytically oriented theories of child development and previous empirical findings, the following hypotheses were formulated for examination in the present research.

Hypothesis I: There is a positive association between the self-concept and the mother concept.

Hypothesis II: There is a negative association between the self-concept and food aversions.

Hypothesis III: There is a negative association between the mother concept and food aversions.

Hypothesis IV: In comparison with disturbed children, normal children have a higher self-concept, higher mother concept, and fewer food aversions.

METHOD

Subjects

There were two groups of subjects in this study. The normal group consisted of 30 boys, ranging in age from 9–2, to 12–9, with a mean age of 11 years. These boys were attending a YMCA summer camp, and were mostly representative of average middle class families with some from the lower classes. The disturbed group consisted of 25 boys, ranging in age from 10–0 to 12–4, with a mean of 11 years. They were undergoing residential psychiatric treatment for various emotional disturbances, and most of these boys came from middle and lower class families.

Measures and Procedure

The following instruments were administered in individual testing sessions:

Adjective check list. An adjective check list was used, consisting of 20 direct questions concerning the self. Each adjective was preceded by the question, "Are you usually _____?" and the child answered "yes" or "no." Ten of the adjectives were positive, or socially desirable (e.g., happy, helpful, popular, etc.), and 10 were negative, or socially undesirable (e.g., selfish, hostile, fearful, etc.). Scoring of responses to this instrument consisted of adding the number of yes responses to the positive adjectives and the number of no responses to the negative adjectives, which provided a measure of positive self-concept. The number of yes responses to the negative adjectives and the number of no responses to the positive adjectives were summed to provide a measure of negative self-concept. With this procedure, the highest positive score could be 20 and the highest negative score could be 20. In order to arrive at a single self-concept score for each child, the negative score was subtracted from the positive score, and this score, which could be either plus or minus, was used in the statistical analyses.

Picture self-rating. Another self-rating device consisted of an $8\frac{1}{2}'' \times 11''$ piece of cardboard on which were pasted 10 small circles. The child was told that the circles were supposed to represent the 10 boys in his tent (camp subjects) or the 10 boys in his living unit (institutionalized subjects). The circles were numbered from 1–10 and, in reference to 8 different traits, the child was asked to indicate his relative position in the group. In other words, for each of the 8 traits, of which 4 were positive (e.g., popular, capable, etc.) and 4 were negative (e.g., aggressive, dependent, etc.), the child was told that Circle Number 1 represented the boy who was lowest on the trait and Circle 10 the boy who was highest on the trait, and he was asked to indicate his relative standing in the group. The child's ratings from 1–10 on each of the positive traits were added to provide a measure of positive self-concept, and the ratings on the negative traits were

[1] The authors wish to express their appreciation to Charles Hanscomb, the director of the YMCA summer camp in Wakefield, Rhode Island, for providing the subjects for the normal group. Gratitude is also acknowledged to Maurice W. Laufer and the staff of the Bradley Hospital in Riverside, Rhode Island, for facilitating the study of the emotionally disturbed children.

[2] Now at Northwestern University.

added to provide a measure of negative self-concept. In order to arrive at a single over-all self-rating, the negative score was subtracted from the positive score, resulting in a final score that could range from a low of −36 to a high of +36.

Food preference inventory. Food preferences were assessed by an extended version of Wallen's (1945) list of foods. The child was asked to indicate his feeling toward each of the 50 foods on the list. The responses were in terms of the following categories: (*a*) like; (*b*) never tasted, but think I would like; (*c*) dislike; (*d*) never tasted, but think I would not like; and (*e*) refuse to eat. Scoring of this inventory simply entailed counting the number of responses in each of the five categories. Scores in categories *a* and *b* were summed to provide a total "like" score, and the sum of the remaining three categories provided a total "dislike" score.

Blacky test. The children were presented with six cards from the Blacky test developed by Blum (1950). These cards, Numbers 1, 4, 7, 8, 9, and 11, were selected on the basis of clinical judgment that they provided scenes likely to elicit material indicative of mother-child relations. In response to each picture, the child was asked to answer a series of questions from the Blacky test inquiry. The answers were scored to provide measures of the following characteristics in the child: aggression, anxiety, dependency, pessimism, security, independence, and optimism. For each stimulus card, each variable was scored along a five-point scale, running from a slight implication of the variable to very strong indication of the variable. The total score for the variables of security, independence, and optimism was computed to provide a measure of positive self-concept, and the total score for aggression, anxiety, dependency, and pessimism served as a measure of negative self-concept. An over-all self-concept score was obtained by subtracting the negative score from the positive score. The answers to the inquiry were also scored to provide indices of the following characteristics in the mother-figure: aggression, anxiety, dominance, rejection, competence, nurturance, and permissiveness. For each stimulus card, each of these variables was scored on a five-point continuum. A positive mother concept score was obtained by adding the scores on competence, nurturance, and permissiveness, and a negative mother concept score resulted from summation of the scores on aggression, anxiety, dominance, and rejection. An over-all mother concept score was obtained by subtracting the negative score from the positive score.

Story completion technique. The Story Completion test was modified from a procedure employed previously by Barndt and Johnson (1955), and Davids and Parenti (1958). Briefly, the procedure consists of presenting the child with a sentence describing some event and asking him to make up a story using the sentence as a stimulus. In the present study, the following five sentences were used to elicit story completions: a lady and her little boy went to visit the zoo one day; a young boy brought a new friend home from school one day to meet his mother; when a certain boy woke up, he knew that something unusual was going to happen today; while a boy and his sister were playing, they knocked over the lamp that their father had given their mother for her birthday; a boy wanted to go out and work in the garden with his daddy, but his mommy wouldn't let him go. This procedure is designed to elicit fantasy material similar to that obtained from TAT stories, but enabled us to present stimuli that seemed likely to call forth fantasies concerning mother-child relations. The thematic content of the stories was scored for the same variables, and employing the same scoring scheme, as with the Blacky test. Thus, an over-all self-concept score and a mother concept score were obtained from this technique.

Reliability of scoring. Two raters independently scored responses to the projective techniques. The interrater agreement, as measured by product-moment correlations, was .95 for the self-concept scores and .75 for the mother concept scores derived from the story completion technique, and .90 and .92 for the self-concept and mother concept scores derived from the Blacky test. These cofficients, which are significant well beyond the .01 level, indicate a quite satisfactory degree of concordance in the scoring of this projective material.

Statistical treatment of the data. Within the two groups, correlations among the self-concept, mother concept, and food aversions scores were computed by the product-moment method. Differences between the two groups on each of these three variables were evaluated by Bartlett's test for homogeneity of variance and *t* tests between means. Since the nature of associations and direction of differences were predicted in advance, one-tailed tests of significance were employed.

RESULTS

Findings within the Normal Group

As shown in Table 1, the results derived from both projective tests indicate significant positive association between measures of the self-concept and the mother concept. Furthermore, the findings presented in Table 2 show a consistent tendency for a low self-concept to be associated with a high number of food dislikes. Of the four measures of self-concept, the one derived from the adjective check list shows a significant negative correlation with food aversions. Moreover, there are two significant correlations between measures of low self-concept and relatively high number of foods that have never been tasted but are

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN SELF-CONCEPT AND MOTHER CONCEPT

Measures	Normal Group (N = 30)	Disturbed Group (N = 25)
Blacky test	.65**	.34*
Story completions	.55**	.50**

* Significant at the .05 level for a one-tailed test.
** Significant at the .01 level for a one-tailed test.

TABLE 2

PRODUCT-MOMENT CORRELATIONS BETWEEN SELF-CONCEPT AND FOOD AVERSIONS IN THE NORMAL GROUP
($N = 30$)

Self-Concept Measures	Food Inventory		
	Total Dislike	Never Tasted, Dislike	Like
Adjective check list	−.41*	−.35*	.82**
Picture ratings	−.15	−.02	.19
Blacky test	−.04	−.83**	.24
Story completions	−.29	−.23	.54**

* Significant at the .05 level for a one-tailed test.
** Significant at the .01 level for a one-tailed test.

disliked. In this regard, the magnitude of the correlation (−.83) between the Blacky test measure of self-concept and the "never tasted, but dislike" responses on the food aversion list is particularly noteworthy. Comparison of the self-concept indices with the number of foods indicated as liked also reveals quite striking associations. Both the adjective check list (objective method) and the story completion technique (projective method) reveal highly significant correlations (.82 and .54) between a high self-concept and liking a relatively great variety of foods. The findings in regard to mother concept and food aversions, contained in Table 3, also show consistent trends in the predicted direction. Moreover, there is a significant positive correlation between high mother concept and number of foods indicated as liked. Thus, the results obtained from the group of normal children are in keeping with the predictions formulated in Hypotheses I, II, and III.

Findings within the Disturbed Group

As shown in Table 1, with measures derived from both projective methods, there is significant positive association between the self-concept and mother concept, with children who have a relatively high self-concept tending to also have a high mother concept. In the disturbed group, the findings shown in Table 4 are generally consistent with the predictions. Low self-concept, measured by the picture self-rating technique, is correlated significantly with both a high total number of disliked foods and high number of foods that have never been tasted but are indicated as disliked. The findings presented in Table 5 are not consistently in accord with the predictions and, in general, do not support Hypothesis III. However, the one statistically significant coefficient indicates a relationship between a low mother concept, defined by the story completion technique, and a high number of foods that have never been tasted but are disliked. Thus, the results obtained from the disturbed children confirm Hypothesis I, lend

TABLE 3

PRODUCT-MOMENT CORRELATIONS BETWEEN MOTHER CONCEPT AND FOOD AVERSIONS IN THE NORMAL GROUP
($N = 30$)

Mother Concept Measures	Food Inventory		
	Total Dislike	Never Tasted, Dislike	Like
Blacky test	−.12	−.21	.55**
Story completions	−.16	−.23	.17

** Significant at the .01 level.

TABLE 4

PRODUCT-MOMENT CORRELATIONS BETWEEN SELF-CONCEPT AND FOOD AVERSIONS IN THE DISTURBED GROUP
($N = 25$)

Self-Concept Measures	Food Inventory		
	Total Dislike	Never Tasted, Dislike	Like
Adjective check list	−.09	−.11	.01
Picture ratings	−.37*	−.54**	.28
Blacky test	−.04	.06	.07
Story completions	.32	.15	−.22

* Significant at the .05 level for a one-tailed test.
** Significant at the .01 level for a one-tailed test.

TABLE 5

PRODUCT-MOMENT CORRELATIONS BETWEEN MOTHER CONCEPT AND FOOD AVERSIONS IN THE DISTURBED GROUP
($N = 25$)

Mother Concept Measures	Food Inventory		
	Total Dislike	Never Tasted, Dislike	Like
Blacky test	.17	.30	−.02
Story completions	.16	−.34*	−.01

* Significant at the .05 level for a one-tailed test.

partial support to Hypothesis II, and fail to confirm Hypothesis III, although they are somewhat equivocal in this last regard.

Comparison of Findings in the Two Groups

In evaluating Hypothesis IV, it is necessary to compare the two groups on scores indicative of self-concept, mother concept, and food aversions. First, the results presented in Table 6 show that, on the four different measures, the normal children have a higher self-concept. Three of these differences are statistically significant well beyond the .01 level. On the picture self-rating procedure, which is the only one not yielding a significant difference between means, Barlett's test indicates that the disturbed children are significantly more variable in the self-concept scores obtained by this direct method of assessment. As shown in Table 7, in response to both projective methods employed in this study, the normal children have a significantly higher mother concept than do the disturbed children.

TABLE 6
COMPARISON OF SELF-CONCEPT SCORES IN THE TWO GROUPS

Measures	Group				F	t
	Normal ($N = 30$)		Disturbed ($N = 25$)			
	M	Variance	M	Variance		
Adjective check list	11.9	29.9	7.4	37.7	1.26	2.94**
Picture ratings	13.5	84.7	11.6	165.7	1.96*	.60
Blacky test	−1.1	11.7	−4.4	6.1	1.91*	4.11**
Story completions	3.6	31.2	−1.2	21.6	1.44	3.53**

* Significant at the .05 level.
** Significant at the .01 level.

TABLE 7
COMPARISON OF MOTHER CONCEPT SCORES IN THE TWO GROUPS

Measures	Group				F	t
	Normal ($N = 30$)		Disturbed ($N = 25$)			
	M	Variance	M	Variance		
Blacky test	.5	2.6	−1.5	2.6	1.00	2.31*
Story completions	.2	5.8	−1.1	7.9	1.36	2.64**

* Significant at the .05 level.
** Significant at the .01 level.

TABLE 8
COMPARISON OF FOOD AVERSION SCORES IN THE TWO GROUPS

Measures	Group				F	t
	Normal ($N = 30$)		Disturbed ($N = 25$)			
	M	Variance	M	Variance		
Total Dislike	10.2	22.2	12.1	40.0	1.80	2.16*
Refuse to Eat	3.3	11.8	5.6	18.8	1.59	2.14*
Never Tasted, Dislike	3.3	4.9	4.1	10.6	2.18*	1.00
Like	37.7	22.4	36.4	53.2	2.38*	.27

* Significant at the .05 level.

The group differences in regard to food aversions, shown in Table 8, are also consistent with the prediction. The two statistically significant differences indicate that, as a group, the disturbed children have more food dislikes and also refuse to eat more foods than do the normal children. The two nonsignificant differences suggest a tendency for the disturbed children to indicate fewer foods as liked and more foods that have never been tasted, but are disliked. In regard to these trends, it is noteworthy that on both measures the disturbed children are significantly more variable than the normal children.

Thus, viewed as a whole, the statistical findings confirm Hypothesis IV.

DISCUSSION

The results obtained in the present investigation lend empirical support to the theoretical and clinical conviction that mother-child relations play a crucial role in the formation of the child's self-concept. The case histories of the emotionally disturbed children who were studied in this research contain considerable evidence of traumatic early experiences, with much turmoil in the child's early home environment. Often the child experienced little emotional warmth and little psychological security. As Erikson (1950) has emphasized, if the child does not experience consistent mothering and does not receive loving gratification of his needs, he is likely to develop a basic mistrust of himself and the world around him. Instead of developing confidence, he becomes anxious, unhappy, and unsure of himself and others. In keeping with

this psychoanalytically oriented theory of ego development, in the present study it was found that emotionally disturbed children, the majority of whom had negative family experiences from very early childhood, had a lower self-concept and a lower mother concept than did a group of normal children. The present findings do not provide answers about cause and effect, but they do show that the more positive the child's self-concept, the more positive is his concept of the mother. That is, the children whose self-picture contains signs of such characteristics as competence, independence, optimism, and social facility, tend to view the mother figure as being competent, nurturant, and emotionally stable.

In the group of normal children, both the self-concept and the mother concept are related to measures of food aversions, with children who have a more positive self-picture and more positive mother concept tending to have relatively few food aversions and greater liking for a variety of foods. In the disturbed group, the self-concept appears to be related to food aversions, with fewer aversions in those children who have a relatively high self-concept. In this group, however, the relationship between mother concept and food aversions is not so clear-cut. In general, the disturbed children evidence a greater number of food dislikes and refuse to eat a significantly greater number of foods than do the normal children. Particularly noteworthy is the finding that the disturbed children indicate a greater number of foods that they have never tasted but feel they would dislike. The disturbed children are also significantly more variable in their scores indicative of food likes, with some of them saying that they like practically every item on the list and others indicating very few foods that they actually like.

At present, there is almost universal agreement among psychologists and psychiatrists that the mother figure, love, and food are interwoven in the child's unconscious processes. There is little doubt that the feeding relationship between the mother and child during the oral stage of psychosexual development is of high importance in the youngster's developing psychodynamics. The way in which the feeding situation and oral gratification are handled or mishandled can have lasting effects on the child's intellectual and emotional development. In the group of institutionalized children, in whom the case records provide ample evidence of early feeding conflicts, there are much more variable reactions to foods, and these reactions do not appear to be systematically related to the child's conception of the mother as evidenced in his fantasies.

If the developmental process has been more orderly, as in the case of the normal children in this research, then relations between the mother concept and food aversions should be more predictable and understandable. With severely disturbed children, on the contrary, the specific types of psychopathology and inappropriate use of defense mechanisms should lead to more complex and equivocal relations that fit less well with predictions based on theories of normal child development.

In the present study, the focus is on statistical relationships and group differences, not on individual cases. However, by way of pointing to some of the complexities and perplexing phenomena that await greater clinical understanding, let us describe briefly two extreme examples from the emotionally disturbed group of boys. One boy who was very obese and actually ate voraciously avowed that he liked an extremely small number of items on the food inventory. Further evincing his preoccupation with foods and oral activities, he was in the habit of calling other people "fat turkey" and would tell various staff members that they ate too much and were too fat. Another boy, however, who was exceptionally underweight and actually looked like the emaciated people who were released from concentration camps, indicated that he liked practically every item on the food inventory. This boy, who was often teased about his physique by the other children and was nicknamed "Boney," had an extremely ambivalent relationship with his mother from infancy into later childhood. She was overprotective yet demanding, seductive yet hostile. Thus, with cases such as these, there were ample grounds for inconsistencies between conscious and unconscious concepts of self and mother, as well as inconsistencies in relations between these schema and avowed feelings about food.

In conclusion, then, it should be emphasized

that in conducting research on such complex subject matter, with extreme degrees of psychopathology, neurotic interpersonal relations, unconscious conflicts, and confusions between fantasy and reality, one should expect ambiguities in experimental findings. However, in the present research, all hypotheses were confirmed with the group of normal children, and findings from the disturbed children were generally consistent with theoretical expectations. Thus, the theoretical framework used here seems fruitful for research on the comparative psychological development of disturbed and normal children.

SUMMARY

Measures of self-concept, mother concept, and food aversions were obtained from a group of 11-year-old normal boys and a group of 11-year-old boys who were institutionalized for emotional disturbances. Indices of self-concepts and mother concepts were derived from direct methods and projective methods of assessment, and food aversions were assessed by an objective inventory. It was hypothesized that self-concepts and mother concepts are positively associated, and there are negative correlations between these measures and number of food aversions. These predictions were confirmed in the normal group and generally supported by findings from the emotionally disturbed group. However, there were more inconsistencies in the relations found in the disturbed sample. As predicted, in comparison with the disturbed children, the normal boys evidenced a significantly higher self-concept, higher mother concept, and fewer food aversions.

Some relations between these empirical findings and theories of child development were discussed, and pathological early mother-child relations were considered in accounting for the findings from the institutionalized children. It was concluded that the results of this study are quite encouraging and that, in spite of the complexities involved in research on such subject matter, much can be learned about the process of child development and personality formation from comparative studies of normal and emotionally disturbed children.

REFERENCES

BARNDT, R. J., & JOHNSON, D. M. Time orientation in delinquents. *J. abnorm. soc. Psychol.*, 1955, **51**, 343–345.

BLUM, G. *The Blacky Pictures*. New York: Psychological Corporation, 1950.

DAVIDS, A., & PARENTI, ANITA N. Time orientation and interpersonal relations of emotionally disturbed and normal children. *J. abnorm. soc. Psychol.*, 1958, **57**, 299–305.

ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.

WALLEN, R. Food aversions of normal and neurotic males. *J. abnorm. soc. Psychol.*, 1945, **40**, 77–81.

(Received February 15, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 315-321

CHILDREN'S ATTITUDES TOWARD RATIONAL VERSUS INHIBITING PARENTAL AUTHORITY

ANATOL PIKAS
Uppsala University, Sweden

ALTHOUGH the work by Adorno, Frenkel-Brunswik, Levinson, and Sanford (1950) has shaped the concept of the authoritarian personality for social psychological research, the book did not define the term, and, as has often been pointed out, it lacks a concise statement of the hypotheses of the research project as a whole (Christie & Jahoda, 1954; Smith, 1950). It has been observed that *The Authoritarian Personality* began as a study of the Fascist personality. The essential traits of the syndrome with which the book deals center around problems of ethnocentrism, and it seems likely that the book was not given its title until the final stage of the research (Christie & Jahoda, 1954). The present author has been especially struck by the fact that in *The Authoritarian Personality* and still more in later research originating in the F Scale, there is little explicit attention to the conception of authority (Gregory, 1955).

Fromm (1941), considered to have introduced the concept of the authoritarian personality in modern psychology (Bay, 1958), based his definition on the word *authority*:

> Authority refers to an interpersonal relation in which one person looks upon another as somebody superior to him. But there is a fundamental difference between a kind of superiority-inferiority relation which can be called rational authority and one which may be described as inhibiting authority.

Both enactment of the role of *inhibiting* authority and the tendency to submit to this type of authority were included by Fromm (1941) in the "authoritarian personality" syndrome that became the subject of extensive treatment. On the other hand, neither enactment of the role of *rational* authority nor attitudes towards such authority were extensively treated by Fromm or by Adorno et al. (1950). The purpose of this study is to re-establish the distinction between the two types. Attitudes of adolescent children towards parental authority are studied with respect to the crucial context in which the authority expresses restrictive norms and performs actions that block or check the spontaneous activity of the adolescent. The basic assumptions underlying the variables and measures of the study are these:

1. Norms and actions of the parents that hinder the spontaneous activity of the adolescent may, according to the kind of justification they are given, be a manifestation of either rational or inhibiting authority.

Definition I. Attitudes of children towards those restrictive norms and checking actions of their parents that are accompanied by motivations giving expression to the parents' aims for the children's welfare, are called the Ara variable ("A" for authority and "ra" for rational); children's attitudes towards those restrictive norms or checking actions of their parents that are accompanied by motivations expressing egoistic domination or the desire for authority for authority's own sake, are called the Ainh variable ("A" for authority and "inh" for inhibiting).

Since measures of the two variables were constructed for this study in such a way that members of each pair of items were identical to each other except for ascribed parental motivations, the variable Ara should elicit more positive attitudes than the variable Ainh (Ara > Ainh). This assumption has an axiomatic rather than a hypothetical nature. If Ara > Ainh holds empirically, the validity of the test is supported.

2. In these two variables a distinction is made between two types of items designated by (c) and (a).

Definition II. Restrictive norms or checking actions that have a relatively *concrete* character, as pertaining to given material objects, are represented by the letter c in parentheses, which is added to the variable symbol. The restrictive norms or checking actions that belong to more *abstract* or general relations are represented by the letter a in parentheses.

Thus a distinction is made between Ara(c) and Ara(a), likewise between Ainh(c) and Ainh(a). The distinction between (c) and (a) was introduced because pilot investigations

showed that restrictive norms and checking actions concerning relatively concrete objects or relations tended to evoke more negative attitudes.

3. In order to stimulate the child's imagination, items to measure the Ara and Ainh variables were given the form of unfinished implication-sentences (if . . . then . . .). All the sentences were given in both direct and projective form. In the latter case the children answered on behalf of an imaginary child of their own age. The variables measured projectively are represented by Pr and those directly obtained by Di. Thus so far eight variables have been defined: Ara(c)Pr, Ara(a)Pr, Ainh(c)Pr, Ainh(a)Pr, Ara(c)Di, Ara(a)Di, Ainh(c)Di, Ainh(a)Di.

Attitudes of children towards the restrictive norms and checking actions of their parents should be more positive in the direct than in the projective form of the test (ADi > APr).

4. Because of methodological interest in results yielded by the incomplete implication sentences and by more conventional attitude questions, two more variables were added to the system.

Definition III. A(c) and A(a) are variables having the same content as Ara and Ainh except that the items have the form of attitude questions and lack the additions showing whether the parental authority is of a rational or inhibiting nature.

The A(c) and A(a) variables occur only in the direct form.

5. To permit exploration of how the variables thus far defined correlate with the subjects' *own* responses in the role of rational or inhibiting authority, a final variable was added.

Definition IV. By variable Ora is meant the extent to which the adolescent prefers altruistic or task centred alternatives to authoritarian ones in imaginary occupational situations.

The O variable was measured both directly and projectively in the same way as the Ara and Ainh variables.

6. The investigation set out to study all the foregoing variables in relation to the chronological age of the subject. The recognized fact that authoritarian scales correlate negatively with intelligence (Titus & Hollander, 1957) makes it important to control for intelligence level. The battery of the tests was therefore supplemented by a short verbal intelligence test. The general line of thought here was that besides the *emotional* preferences that the attitude inventories were to measure, one must also reckon with *cognitive* factors. On the view that the intellectually less developed person is, in Lewin's terms, not well structured cognitively, he should show a tendency to agree with the authoritarian items because these provide a simpler solution. According to this reasoning two basic assumptions are advanced which can also be considered as the chief hypotheses of the study:

a. Younger children, as cognitively less well structured than older children, show more approval than do older children of parental checking actions that have purely *authoritarian* motivations: Ainh *decreases* with age.

b. Variables that reflect the function of *rational authority* show the opposite relation: Ara *increases* with age.

METHOD

Subjects

The study used 656 boys and girls from four different schools: two in Stockholm, one in Uppsala, and one from a community in central Sweden. The schools differed in social class and standard of intellectual ability. In order to obtain an equal number from each social group and sex on each of the five age levels, a stratified sample of 300 subjects was selected: 10 boys and 10 girls from each of the three social classes on each age level. The results pertain partly to the total number of subjects, partly to the stratified sample.

Instruments

Three types of inventories were used, as summarized in Table 1. Examples of items assigned to particular variables follow.

Items in the two A inventories have certain similarities in content. Two items from the A inventory of attitude questions are:

Item 45 in A(c)
Should children go to bed at a set time? YES ? NO
Item 52 in A(a)
Do you think that parents should sometimes be strict? YES ? NO

The A(c) item has two corresponding items in the boys' projective version of the implication sentence inventory:

Item 1 in Ara(c)Pr
If his parents think that he should go to bed at a set time in order to get enough sleep, he thinks that
Item 8 in Ainh (c)Pr
If his parents think that he should go to bed at a set time so that they can be left in peace, he thinks that

TABLE 1

VARIABLES MEASURED BY THE THREE INVENTORIES

Nature of Inventory	Symbols of the Variables	Description of Elements Within the Variables		
		Description of the Dimensions	Division of Items into (c) and (a) and description of Individual Variables	
			Concrete Items	Abstract Items
A inventory consisting of implication sentences	Ara	Parental checking action as an expression of *rational authority*	Ara items of a more *concrete* nature = Ara(c)	Ara items of a more *abstract* nature = Ara(a)
	Ainh	Parental checking action as an expression of *inhibiting authority* (authoritarian motives)	Ainh items of a more *concrete* nature = Ainh(c)	Ainh items of a more *abstract* nature = Ainh(a)
A inventory consisting of attitude questions	Dimensions ra and inh not present in this inventory		Parental restrictive norms of a more *concrete* nature = A(c)	Parental restrictive norms of a more *abstract* nature = A(a)
O inventory consisting of paired alternatives representing imagined occupational situations	Ora	Alternatives that reflect altruistic or task centered motives	Dimensions (c) and (a) are not present in the O inventory	
	Oinh	Alternatives that reflect authoritarian motives		

Note.—All the variables with the exception of A(c) and A(a) were tested both projectively (Pr) and directly (Di). Each of the variables thus obtained has nine items.

The A(a) item has in its turn two corresponding items in the implication sentence inventory:

Item 4 in Ara(a)Pr

If his parents are strict sometimes, because they want him to grow up to be a good person, he thinks that

Item 23 in Ainh(a)Pr

If his parents are sometimes strict because they like to decide things for him, he thinks that

The direct form of the implication sentence inventory differed from the projective form in that the unfinished sentences were constructed in the first person singular instead of in the third person singular. For example:

Item 1 in Ara(c)Di

If my parents think that I should go to bed at a set time in order to get enough sleep, I think that

The subjects' responses to the implication sentence inventory were classified by seven independent raters, two for each subject, according to the rotation principle.

Items in the O inventory involved a forced choice between altruistic and authoritarian alternatives. Here are two examples from the Pr form for boys:

As a judge, he feels most satisfied
1. when he knows that no criminal escapes punishment.
2. when all verdicts are as just as possible.

As the captain of a large merchant ship, he feels most satisfied
1. when the ship carries its cargo quickly to the port of destination.
2. when he is allowed to command a large crew alone.

In the girls' version typical male occupations were replaced by female ones, the status and work corresponding as far as possible in relevant aspects. Preference for the altruistic or task centred alternative is counted as the Ora variable.

The inventories were administrated in the following order:

1. The A inventory consisting of implication sentences in projective form. Subjects were instructed that "here is an imagination test" and that they were to "imagine a boy (girl), give him (her) a name they liked," and "fill in sentences just as he (she) would have done." The four different variables within the inventory were mixed at random.
2. The A inventory consisting of attitude questions with A(c) and A(a) items mixed at random.
3. The projective form of the O inventory.
4. The verbal intelligence test.
5. The direct form of the A inventory consisting of implication sentences.
6. The direct form of the O inventory.

Completing the inventories occupied two consecutive class periods of 45 minutes each.

RESULTS

Reliability of Raters

Responses to the incomplete sentences comprising the Ara and Ainh measures were classified as reflecting positive and negative attitudes with satisfactory reliability. Only 2% of the answers could not be classified (for only

one item was the value as high as 5%) and the phi coefficient between independent raters was greater than .95.

Reliability and Homogeneity

In 60 different item analyses (12 variables at five different age levels each), the difference between the higher and lower halves was strongly in the expected direction for each item, with two exceptions: one item in Ara-(c)Pr showed a reversed direction in four of the age levels, and two other items in Ara-(c)Pr and Ara(a)Di showed no difference between the higher and lower halves in their respective age levels.

The Kuder-Richardson reliability, calculated for each variable in the age levels 11, 13, and 15, was generally low, one reason being the shortness of the nine-item inventories. The average for the 36 K-R coefficients was .54.

Construct Validity

The assumption that the subjects would be more positive towards parental restrictive norms and checking actions when these reflect the parental role as rational authority (Ara) than when they reflect purely authoritarian parental motives (Ainh) is empirically confirmed ($p < .005$). Similarly, the assumption that more positive attitudes towards the restrictive norms and checking actions of parents would be elicited by direct than by projective items was confirmed ($p < .005$) for the data from the entire sample.

Developmental Trends

The results for the selected sample equated for social class did not differ importantly from those for the total sample with respect to age trends. Only the general trends for the equated sample are considered here. The curves presented in Figures 1–4 warrant the following general conclusions:

1. All the Ainh variables decrease. In other words, positive attitudes towards parental restrictive norms and checking actions dictated by authoritarian motives decrease with age (Figure 1).

2. The Ara variables do not show uniform trends. When trends appear, namely, in Ara(c)Di and Ara(a)Pr they have an opposite tendency; (*a*) Ara(c)Di decreases and (*b*) Ara(a)Pr increases (Figure 2). Or to put it another way: (*a*) Positive attitudes towards parental restrictive norms and checking actions, accompanied by *rational* motivations, *decrease* with age when the items are of a more concrete nature and the attitudes have been expressed in the *direct* form of the schedules; (*b*) Positive attitudes towards parental restrictive norms and checking actions, accompanied by *rational* motivations, *increase* with age when the items are of a more abstract nature and the attitudes have been expressed in the *projective* form of the schedules.

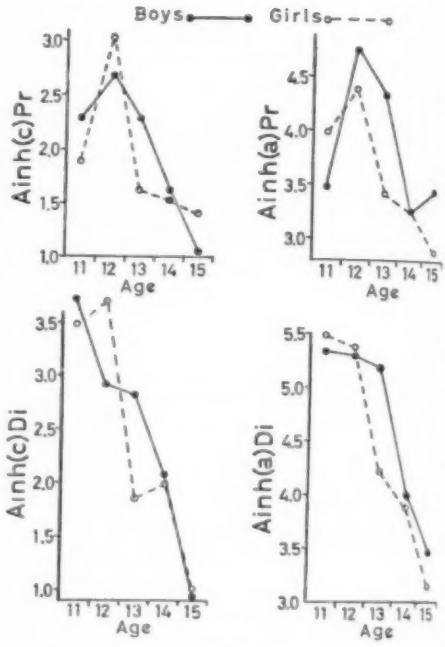

FIG. 1. Developmental trends of the four Ainh variables.

3. The A variables consisting of simple attitude questions show a tendency to decrease, which is more clearly marked in A(c) than in A(a). That is, positive attitudes towards general restrictive norms of upbringing *decrease* with age, this tendency being more obvious in items of a more *concrete* content (Figure 3).

4. The variable Ora shows a tendency to rise, that is to say, preference for altruistic alternatives over authoritarian ones in imaginary occupational situations tends to become greater (Figure 4). Ora is the only variable

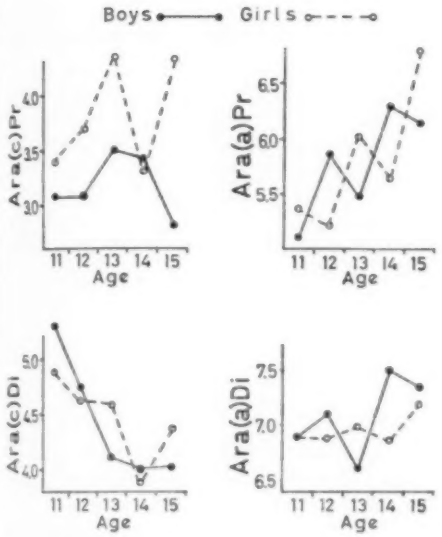

FIG. 2. Developmental trends of the four Ara variables.

that shows distinct sex-differences. However, since the boys' or girls' inventories contained different occupations the frequency of the answers of the sexes cannot be compared.

Correlation of Attitude Variables with Verbal Intelligence and with Preferences for Altruistic or Authoritarian Alternatives (O)

The correlational data of the study, comprising tetrachoric coefficients based on dichotomizing the data as nearly as possible to the median, are presented in Table 2. The correlations with verbal intelligence and with

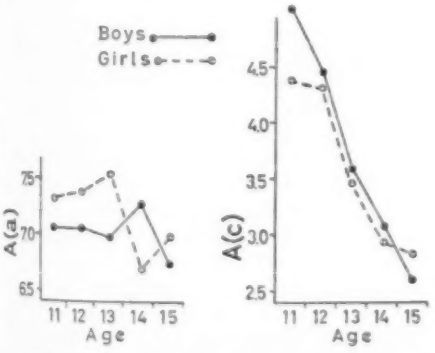

FIG. 3. Developmental trends of the two A variables.

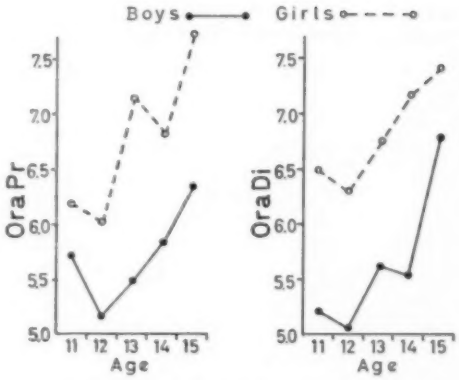

FIG. 4. Developmental trends of the two Ora variables.

Ora are based on the smaller stratified sample. Correlations of the attitude variables with the chronological age (CA) are based on the entire larger sample.

A study of the columns of the table for Ora correlations shows that although the individual correlation coefficients seldom attain significance, when regarded as a whole they exhibit a pattern that appears to justify two general conclusions: Ora correlates positively with Ara; and Ora correlates negatively with Ainh. One of the Ara variables is an exception to this, namely, Ara(c)Di which correlates negatively with Ora on one age level. In other words, children's preferences for *altruistic* occupational motives as opposed to authoritarian ones are positively related to their attitudes toward rationally motivated checking actions by their parents, and negatively related to their attitudes toward authoritarianly motivated parental checking actions.

Regarding the connection between the attitude variables and verbal intelligence, the preference for *altruistic* alternatives (rather than for authoritarian ones) in imaginary occupational situations (Ora) increases with rising verbal intelligence.

In regard to the correlation of the A variables with verbal intelligence, the general trend of the relationship agrees with the age trends already described. Ainh shows on the whole negative correlations, while the pattern of Ara correlations is not so clear. Because of the incompleteness of the table of correlations, it would be hard to draw more precise conclusions.

TABLE 2

CORRELATIONS OF ATTITUDE VARIABLES WITH CHRONOLOGICAL AGE (CA), VERBAL INTELLIGENCE, AND THE ORA VARIABLE

Content of Variable	Symbol	CA ($N = 656$)	Verbal Intelligence			Ora Variable[a]		
			Age Groups			Age Groups		
			11 ($N=60$)	13 ($N=60$)	15 ($N=60$)	11 ($N=60$)	13 ($N=60$)	15 ($N=60$)
Conventional attitude questions (concrete)	A(c)	−.49**	−.39*	−.45*	.14	−.45*	−.27	.16
Parental authoritarian checking (concrete and direct)	Ainh(c)Di	−.40**	−.38*	[b]			−.21	
Parental authoritarian checking (abstract and direct)	Ainh(a)Di	−.38**	−.27	−.60*	−.27	−.18	−.51*	
Parental rational authority checking (concrete and direct)	Ara(c)Di	−.21**		−.45*			−.43*	
Parental authoritarian checking (concrete and projective)	Ainh(c)Pr	−.18**	−.13	−.50*	−.28	−.15	−.28	
Parental authoritarian checking (abstract and projective)	Ainh(a)Pr	−.14*	−.26	−.20		−.21	−.33*	−.38*
Conventional attitude questions (abstract)	A(a)	−.13*	.29				.28	
Parental rational authority checking (concrete and projective)	Ara(c)Pr	.02			.40*	.34*	.26	
Parental rational authority checking (abstract and direct)	Ara(a)Di	.07	.31	.24	.23	.28	.20	
Parental rational authority checking (abstract and projective)	Ara(a)Pr	.14*	.15			.38*	.23	
Preference for altruistic motives to authoritarian (projective)	OraPr	.17**	.39*	.76*	.32*	—	—	—
Preference for altruistic motives to authoritarian (direct)	OraDi	.30**	.26	.55*	.17	—	—	—

[a] ADi variables have been correlated with OraDi; APr variables have been correlated with OraPr.
[b] Where the scatter diagrams failed to satisfy requirements of linearity or homoscedasticity, no coefficient is presented.
* Significant at the .05 level.
** Significant at the .01 level.

DISCUSSION

As the present study seems to be the first empirical investigation to stress the distinction between rational and inhibiting authority, it has been difficult to suggest a precise system of hypotheses. The general hypothesis advanced at the outset was that the variable Ainh, which could be considered to measure authoritarian attitudes, would decrease, and Ara, or positive attitudes towards rational authority, increase during the developmental stage of 11–15 years. In the same way, Ora—preference of altruistic to authoritarian alternatives in imaginary occupational situations—would increase.

This approach may be characterized as cognitive and formalistic. The distinction drawn between the two types of authority is based on the parental motives to which the parental restrictive norms and checking actions are ascribed in the situational questions presented to the children. Certain cognitive qualities are required of the subjects before they can at all understand the rational motives accompanying the checking actions and give the appropriate answer. According to this view, the positive correlations of attitude variables with verbal intelligence should be regarded as the primary relationship and those with chronological age as secondary: the older chil-

dren give certain answers only because of their increased intellectual ability.

Besides the cognitive factors, the investigation also involves variables that may be identified as conative, in the distinction between (c) and (a) elements. The (c) items, where the parental checking actions are given a concrete form, evoke, according to our assumptions, more ego involvement than the (a) items, which are more general and abstract. We have seen that the (c) elements contribute to the decrement and the (a) elements to the increment of the positive attitudes toward parental checking actions and restrictive norms during the developmental stage of 11–15. These findings may be understandable in terms of ego development. With increasing age, the young adolescent has less need for the support of parental norms concerning concrete things, and parental restrictions of the parents therefore become more and more threatening to personal freedom.

Similar conative factors may be assumed to be at work in the distinction between direct and projective items. As already noted, the relation ADi > APr is valid for the data as a whole, i.e., children's attitudes towards parental checking actions are shown to be more positive in the direct form. However, as the figures show, this relation is subjected to a characteristic change so that ADi > APr tends with age towards ADi = APr. The decrease of Di with age may be interpreted along lines similar to those invoked with (c): when the items are given in the Di form, older subjects feel more threatened by the restrictions than do younger ones.

SUMMARY

Starting with Fromm's (1941) distinction between rational and inhibiting authority, the study investigates children's attitudes toward their parents' restrictive norms and checking actions, assigned to either rational or inhibiting authority according to their ascribed motivation. Further distinctions between checking actions of a more concrete versus a more abstract nature, between direct and projective forms of inventories, and between different formulations of the items as implication sentences versus conventional attitude questions make a coherent system of variables that is applied according to the cross-sectional method to a sample of 656 Swedish children from 11–15 years of age.

The results indicate that if parents' restrictive norms and checking actions are accompanied by *authoritarian* motivations (inhibiting authority), the children's positive attitudes towards these are negatively correlated with chronological age, verbal intelligence with the age factor eliminated, and preference for altruistic alternatives to authoritarian ones in imaginary occupational situations, but that these correlations are greater for items concerning more concrete situations, and for inventories applied in a direct rather than projective form.

If the restrictive norms and checking actions are accompanied by *rational* motivations (indicating rational parental authority), the attitudes have a tendency to correlate positively with these three independent variables if they are given general and abstract formulations (as opposed to concrete) and the projective forms of inventories are employed.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BAY, C. *The structure of freedom*. Stanford: Univer. Press, 1958.

CHRISTIE, R., & JAHODA, MARIA. (Eds.) *Studies in the scope and method of "The authoritarian personality."* Glencoe, Ill.: Free Press, 1954.

FROMM, E. *Escape from freedom*. New York: Farrar & Rinehart, 1941.

GREGORY, W. E. Authoritarianism and authority. *J. abnorm. soc. Psychol.*, 1955, **51**, 641–643.

SMITH, M. B. Review of T. W. Adorno, Else Frankel-Brunswik, D. J. Levinson, and R. N. Sanford, *The authoritarian personality. J. abnorm. soc. Psychol.*, 1950, **45**, 775–779.

TITUS, H. E., & HOLLANDER, E. P. The California F Scale in psychological research: 1950–1955. *Psychol. Bull.*, 1957, **54**, 47–64.

(Received February 15, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 322-326

PAIRED-ASSOCIATE LEARNING BY SCHIZOPHRENIC AND NORMAL SUBJECTS UNDER CONDITIONS OF PERSONAL AND IMPERSONAL REWARD AND PUNISHMENT[1]

RITA L. ATKINSON
Los Angeles, California

AND

NANCY M. ROBINSON[2]
Chapel Hill, North Carolina

SEVERAL studies have suggested that, for schizophrenics, the motivation to avoid punishment is stronger than the motivation to achieve reward, at least when both reinforcers are mild. When punishment follows any incorrect response or failure to respond, this differential leads to faster learning by schizophrenic subjects under punishment conditions than reward conditions, whereas the reverse is usually true for normal subjects (Bleke, 1955; Buss, Braden, Orgel, & Buss, 1956; Cohen, 1956; Leventhal, 1959; Ullman & Forsythe, 1959). Only one relevant study reports the negative effects of punishment to be stronger than the positive effects of reward with normal subjects (Ferguson & Buss, 1959).

The two separate studies reported here (Atkinson, 1957; Robinson, 1957) were designed to test the generality of these findings with independent samples and to investigate the role of the interpersonal element in the reinforcement. The present experiments differed from previous studies in that the learning task was designed to provide the correct answer mechanically so that the subject was not dependent on the experimenter for information during the learning process, and a control group was employed whose responses received no reinforcement of any kind.

A number of authors have stressed the importance of threats of punishment and recurrent failure in the development of the schizophrenic syndrome. Dollard and Miller (1950), Mowrer (1950), and others have discussed neurotic and psychotic symptoms as habits developed through reduction of a fear drive. For them, escape from unpleasant or punishing situations serves to reinforce the apparently maladaptive behavior characteristic of psychotic individuals. Reward is seen as a weak reinforcer for these individuals because it has represented a conflictful response from the mother and cannot be "trusted." The response to reward may take either the form of blocking or selective inattention. An excellent extension and elucidation of a similar viewpoint has been presented by Rodnick and Garmezy (1957); in particular, these authors give evidence for the importance of the schizophrenic's early relationship with his parents in determining his later learning behavior.

The theoretical framework employed here led to the following predictions:

1. Compared with normal subjects, schizophrenics are more highly motivated to avoid punishments and less highly motivated to attain rewards. Thus, schizophrenic subjects learn more rapidly under conditions of punishment for incorrect responses and for failures to respond than under conditions of reward for correct responses. For normal subjects, the reverse situation holds; learning is faster under reward conditions.

2. Schizophrenic subjects whose incorrect responses and failures-to-respond are punished learn faster than nonreinforced control subjects.

3. Schizophrenic subjects whose correct responses are rewarded learn no faster than controls and may, indeed, learn more slowly.

4. A social reinforcement leads to intensification of the effects exhibited by a mechanical (impersonal) reinforcement.

[1] This report is based upon PhD dissertations submitted in 1957 by Atkinson to the Department of Psychology, Indiana University, and by Robinson to the Department of Psychology, Stanford University. The experiments were conducted while the authors were trainees in the Clinical Psychology Service, Veterans Administration Hospital at Palo Alto, California.

[2] Special thanks are due Paul McReynolds, Chief of Research, Psychology Service at that hospital for his invaluable aid and encouragement, and to the many members of the staffs of the Veterans Administration Hospital and nearby Agnews State Hospital who provided facilities and aided in the selection of subjects. The authors are also grateful for the assistance of their respective dissertation committees, D. C. Beier, J. P. Egan, A. M. Binder, and D. Williams at Indiana University, and C. L. Winder, Q. McNemar, and P. W. McReynolds at Stanford University.

METHOD

The first experiment used four groups of 20 female subjects each, a total of 80. Two groups were selected from the female schizophrenic population at Agnews State Hospital to represent individuals in rather acute stages of illness. No subjects had had more than two previous hospital admissions or one year of total hospitalization, or had any received EST or tranquillizing drugs during the 4 months previous to testing. The other two groups consisted of normal female volunteers at the Palo Alto Veterans Administration Hospital.

The second experiment used a total of 100 male schizophrenic subjects, randomly divided into five groups of 20, all drawn from the patient population of the Palo Alto Veterans Administration Hospital and representing those patients who were relatively more acute and in better contact than the average for that hospital. This sample was similar in age to the patient population in the first experiment (means of 32 years in each) but had an average of three hospitalizations, a longer period of continuous hospitalization, and less education (means of 12.8 and 11.6 school years, respectively). Approximately three-quarters of the male group were receiving tranquillizing medication, although none had had EST during the previous 6 months. The normal female volunteers on the average were older (40 years) and better educated (13.9 school years). No subgroup differed significantly from the mean of its respective sample in any of these factors.

The experimental task consisted of learning three lists of 15 pairs of fairly similar adjectives, a pretest list and two test lists. The lists had been equated for difficulty in a pilot study. In both experiments, List I (the pretest list) was the same for all subjects. In the first experiment, the master list of 30 pairs comprising Lists II and III (the test lists) was arranged to provide two lists of 15 pairs each. This master list was circular and was "rotated" with every third item designated as a possible starting position. Each subject was randomly assigned to one of the 10 starting positions with the restriction that each position was used equally often for a given experimental group. This complex procedure, originally intended for the study of serial effects, was abandoned in the second experiment in the interest of simplicity. The 30-item list was split into Lists II and III at a point which had provided equality of difficulty in the first study.

Testing was done individually, the subject being seated in front of a Hull-type memory drum with the experimenter beside him. The drum revolved one place every 4 seconds. The first time through each list, the word-pairs appeared in the window, and the subject was required to say them aloud. During subsequent trials, only the stimulus word appeared, followed in 4 seconds by the pair, then by the next stimulus word, etc. The subject was required to respond before the pair appeared. In other words, the subject was informed by means of the printed words whether or not he had given the correct response; he did not need the experimenter to inform him. Each list was learned to the criterion of one errorless trial. During the first list, no group received any reinforcement. Lists II and III were run under reward, punishment, or control conditions, depending on the subject's group assignment. List I, thus, served as a base line against which the effects of the reward or punishment conditions could be subsequently evaluated. Between Lists I and II and between Lists II and III, the subjects sorted a series of cartoons according to their judgment of the degree of humor involved. This was done primarily to provide a rest period and to minimize interference between lists.

In each case in which a form of "reward" served as the reinforcement, it followed correct responses and was administered as the drum turned to reveal the correct pair. In each case in which "punishment" served as the reinforcement, it followed either incorrect responses or failures to respond, and it, too, occurred as the drum turned to reveal the correct pair.

The samples were divided as follows: one female schizophrenic group, one male schizophrenic group, and one female normal group were designated "personal reward" groups. Likewise, one female schizophrenic, one male schizophrenic, and one female normal group were designated "personal punishment" groups. One male schizophrenic group received "impersonal reward" and one male schizophrenic group received "impersonal punishment." The final male schizophrenic group received no reinforcement at all.

Under personal reward conditions, one of the following statements was made by the experimenter after a correct response to Lists II and III: "good," "umhmm," "right," and "that's right," in random order. Similarly, under personal punishment conditions, following an incorrect response or a failure to respond, the experimenter said, "no," "wrong," "that's wrong," or "that's an error" in an even, quiet tone. No mention was made by the experimenter of the reinforcement.

Subjects in the impersonal reinforcement groups were told at the beginning of List II that the procedure would be the same as on the first list except that the checking system would differ. An old-style Dalton adding machine was employed which made a loud sound when the lever was pulled. The reinforcement, the loud sound, was identical for these impersonal-reinforcement groups except that for the reward group it followed correct responses and for the punishment group it followed incorrect responses or failures to respond. It was made as the drum turned to reveal the correct pair.

At the completion of the session, which averaged an hour in length, the male subjects were asked a series of questions designed to elicit their awareness of the reinforcement, their emphasis on the "rightness" or "wrongness" of their responses, and their conception of the experimenter's role.

RESULTS

Analysis of the results was carried out in terms of trials and errors on each list, and the differences between the subjects' performance on List I and List III (change scores). In general, the data for trials and errors were quite consistent, although differences between groups tended to be cumulative so that the

List II data were not significant. For the sake of brevity, only trials will be reported.

For the female subjects, complex analysis of variance applied to the change scores produced an F ratio of 16.05 for the interaction term, while there were no statistically significant group or treatment effects. This F ratio, significant beyond the .001 level, indicates that the reward and punishment treatments had differential effects on the two diagnostic groups; the normal subjects improved under the reward condition and not under punishment, whereas the schizophrenic subjects showed an increase in trials to criterion under reward and a decrease under punishment.

Simple analysis of variance technique applied to the change data of the male groups yielded an F ratio of 4.04, significant beyond the .01 level. To compare each male group with every other male group, probabilities were calculated on the basis of one-tailed t tests except for those comparisons for which the direction of difference had not been predicted as in the control group against each reward group. Table 1 presents the probabilities, thus, obtained.

Table 1 indicates that the two punishment groups were clearly and significantly different from the two reward groups and the control group, but very similar to each other. The difference between the reward groups and the control group failed to reach the .05 level of significance, however, and the impersonal reward group differed very little from the control group in mean change scores. The difference between the two reward groups failed to reach the .05 level of significance.

TABLE 1

MEAN TRIALS PER LIST FOR EACH EXPERIMENTAL GROUP

(N = 20 per group)

Treatment	Trials			
	List I	List II	List III	List I-List III
Female schizophrenics				
Personal reward	6.95	7.15	7.75	−.80
Personal punishment	6.65	6.10	5.05	1.60
Female normals				
Personal reward	6.35	5.95	5.05	1.30
Personal punishment	5.70	5.40	5.70	.00
Male schizophrenics				
Personal reward	7.50	6.55	9.65	−2.15
Impersonal reward	8.35	6.85	8.70	−.35
No reinforcement	8.30	7.15	8.25	.05
Impersonal punishment	8.80	6.70	7.20	1.60
Personal punishment	8.20	6.95	6.85	1.35

DISCUSSION AND CONCLUSIONS

The results suggest the following conclusions, especially when fortified by further analysis of the kinds of errors made by the subjects and the data from the "awareness" questionnaire administered following List III.

1. Learning by schizophrenic subjects in this kind of experimental situation is more rapid under conditions of mild punishment (personal or impersonal) for incorrect responses and failures to respond than under conditions of mild reward (personal or impersonal) for correct responses. For normal subjects the reverse situation is true; learning is more rapid under reward conditions.

2. In comparison with a procedure which furnishes no reinforcement, punishment actually has a facilitating effect upon learning by schizophrenic subjects. Analysis of the types of errors suggests that the punishment may have facilitated performance by serving to break perseverative tendencies.

3. Reward with schizophrenic subjects in this kind of situation does not function as a positive reinforcement in the sense of increasing the probability of the occurrence of correct responses. This is in opposition to the results found with normal subjects.

4. The situation with respect to the possible decremental effects of reward is less clear. The t values were of such magnitude that it is not possible to say definitely whether the reward groups were clearly different from the control group, in which case the reward may have blocked the learning process in some manner, or similar to the control group, in which case it is probable that the subjects were inattentive to the reinforcement. The possibility remains that with larger groups of subjects, reliable differences would have become apparent.

5. Concerning the personal element in the reinforcement, it is clear from the data that the two punished groups did not differ in change scores. In addition, these two groups showed no significant difference in any of the three categories included in the questionnaire on awareness of reinforcement. Thus, it ap-

pears likely that the two forms of punishment produce very similar effects.

6. The two reward groups, on the other hand, differed in their change scores, with the personal reward group showing greater improvement in learning from List I to List III than the impersonal reward group; however, the difference did not reach statistical significance. On the awareness questionnaire the personally reinforced group emphasized the factor of "being right" more frequently than did the impersonally reinforced group, and much more frequently verbalized the experimenter's role as that of an evaluator.

It appears possible that in this situation, responses to punishment generalized to other impersonal cues but responses to reward did not, the two responses differing markedly in strength. The argument supporting this hypothesis would state that personal (parental) punishment, a clear-cut and strong stimulus, motivates a strong learned "fear" drive, a response which is called forth readily by other "punishing" cues. The schizophrenic is a person who has not learned to differentiate the danger inherent in environmental cues from that inherent in rejection by his parents, and who has, thus, withdrawn as best he can from all possible punishment in his environment. The difference between the two kinds of reward would be expected to appear because reward has not been learned as a strong motivator, and the generalization of effects from one kind of reward to another is not so marked.

An additional word is in order regarding the apparent contradiction between the present studies and those which have employed less specific kinds of reinforcement with schizophrenic subjects. The latter have consistently found that punishment leads to a decrement in performance, accompanied by various kinds of task irrelevant behavior. Persons involved with the care of the mentally ill have consistently advocated therapeutic acceptance and kindly attention toward psychotics, at the same time discouraging punishment and failure experiences. The results of these experiments should not be interpreted as challenging this point of view. In the first place, this experimental situation was specialized in that every incorrect response or failure to respond was punished. In the usual circumstance, a wide range of maladaptive responses is possible, only part of which will be overtly disapproved as "socially unacceptable." In the typical experimental situation in which the subject is told that his previous performance was in general "poor," he may withdraw from the situation or make task irrelevant responses which decrease his vulnerability to punishment but also produce a decrement in performance.

In the second place, hospitalization is not aimed at producing rapid learning, as such, but rather with getting people well. Perhaps "getting well" for the schizophrenic involves reorientation with respect to those reinforcers which are effective and his developing the ability to seek rewards and avoid punishment in a constructive manner.

SUMMARY

Two separate paired-associate learning studies, using a total of 180 subjects drawn from male and female schizophrenic and normal populations, suggest that:

1. For schizophrenic subjects, unlike normal subjects, the motivation to avoid mild punishment is stronger than the motivation to achieve mild reward, so that in a situation in which giving a correct response is the only way to avoid punishment, punishment has a facilitating effect on performance.

2. Reward with schizophrenic subjects in this kind of situation does not serve as positive reinforcement and may actually have some debilitating effects.

3. The two kinds of punishing reinforcement, both with and without the personal element, have similar effects; the two kinds of rewarding reinforcement were not so clearly similar. Differences in the generalization effects of reward and punishment may be involved.

REFERENCES

ATKINSON, RITA L. Paired-associate learning by schizophrenic and normal subjects under conditions of verbal reward and verbal punishment. Unpublished doctoral dissertation, Indiana University, 1957.

BLEKE, R. Interfering response sets as determining factors of reminiscence effects in schizophrenic subjects. *J. Pers.*, 1955, **23**, 479–498.

BUSS, A. H., BRADEN, W., ORGEL, A., & BUSS, E. H. Acquisition and extinction with different verbal reinforcement combinations. *J. exp. Psychol.*, 1956, **52**, 288–295.

COHEN, B. D. Motivation and performance in schizophrenia. *J. abnorm. soc. Psychol.*, 1956, **52**, 186–190.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

FERGUSON, E. L., & BUSS, A. H. Supplementary report: Acquisition, extinction, and counter-conditioning with different verbal reinforcement conditions. *J. exp. Psychol.*, 1959, **58**, 94–95.

LEVENTHAL, A. M. The effects of diagnostic category and reinforcer on learning without awareness. *J. abnorm. soc. Psychol.*, 1959, **59**, 162–166.

MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.

ROBINSON, NANCY M. Paired-associate learning by schizophrenic subjects under conditions of personal and impersonal reward and punishment. Unpublished doctoral dissertation, Stanford University, 1957.

RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1957. Pp. 109–184.

ULLMAN, L., & FORSYTHE, R. P. Responses by normal and schizophrenic subjects under positive and negative examiner reinforcement in a probability learning situation. *Amer. Psychologist*, 1959, **14**, 407–408.

(Received February 19, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 327-337

THE RELATIVE EFFICACY OF VARIOUS TYPES OF PRIOR BELIEF-DEFENSE IN PRODUCING IMMUNITY AGAINST PERSUASION[1]

WILLIAM J. McGUIRE
Department of Social Psychology, Columbia University
AND DEMETRIOS PAPAGEORGIS
University of Illinois

THE disconcerting vulnerability of people's convictions in forced exposure situations provides a provocative contrast to the relative ineffectiveness of the voluntary exposure persuasion campaigns conducted over the mass media (Hovland, 1959). The high effectiveness of forced exposure situations is seen in laboratory experiments, face-to-face discussion, and political indoctrination of captive audiences (United Kingdom Ministry of Defense, 1955; United States Senate, 1956). The various election studies (Berelson, Lazarsfeld, & McPhee, 1954), on the other hand, indicate the relatively small effects of indoctrination campaigns depending on the voluntary exposure of the audience. The student of persuasive communication is inclined to attribute this contrast, at least in part, to the fact that people characteristically defend their convictions by avoiding exposure to counterarguments. This postulate of self-selected exposure to arguments so as to avoid receipt of information at variance with one's already formed opinions has been too widely discussed in the recent literature to need development here (Festinger, 1957; Janis, 1957; Klapper, 1957).

While defense-by-avoidance is likely to be highly effective for belief maintenance so long as the person can adequately regulate his own exposure to arguments, it has the disadvantage of leaving him poorly prepared to resist counterarguments should he be involuntarily exposed to them. Living in an ideologically monolithic environment, the person tends to underestimate the vulnerability of his beliefs and the likelihood of their being attacked. Hence, he will have had little motivation or practice in developing supporting arguments to bolster his belief or in preparing refutations for the unsuspected counterarguments. This postulate of defense-by-avoidance, and its implication of a deficiency in both practice and motivation for developing belief resistance, suggested several hypotheses regarding the immunizing effectiveness of alternative types of prior belief-defense, before forced exposure to massive doses of counterarguments.

The first of these hypotheses is suggested by the closely analogous health problem of the person who has been brought up in so aseptic an environment that he has failed to develop resistance to infection and, hence, although appearing in very good health, proves quite vulnerable when suddenly exposed to a massive dose of an infectious virus. The disease resistance of such a person might be raised by either of two procedures: he might be given supportive therapy—good diet, exercise, rest, etc.—designed to better his physical condition; or alternatively, he might be given an inoculation of the infectious virus itself (in a weakened form) such as would stimulate, without overcoming, his defenses. With respect to developing immunity to specific diseases, the inoculation procedure is generally more effective.

Pursuing this medical analogy into the area of resistance to persuasion, it is hypothesized that the "supportive therapy" approach of pre-exposing a person to arguments in support of his belief has less immunizing effectiveness than the "inoculation" procedure of pre-exposing him to weakened, defense stimulating forms of the counterarguments. This prediction is made primarily for the case of widely accepted cultural beliefs, since the person's inclination to avoid exposure to counterarguments is likely to have been successful with such uncontroverted beliefs. However, even with regard to a culturally controversial issue,

[1] This study is one of a series on developing resistance to persuasion which are being supported by grants from the University Research Board (University of Illinois) and the National Science Foundation.

H. C. Triandis and C. E. Osgood made many helpful suggestions after reading an earlier version of this report.

it would be predicted on the basis of the postulate of defense-by-avoidance that a person would tend to choose his associates so as to remain within an ideologically homogeneous subpopulation. Studies of political behavior (Berelson et al., 1954) demonstrate this social homogeneity even with respect to so controversial an issue as presidential preference during the quadrennial United States campaigns.

The hypothesized superiority of pre-exposure to weakened counterarguments runs counter to recommendations frequently repeated in investigations of prisoner-of-war behavior during the Korean conflict (Kinkead, 1959; United States Senate, 1956). The readiness of American personnel under prison camp conditions to surrender their prior convictions regarding proper behavior has been attributed to the failure of our formal and informal instructional institutions—home, school, army, etc.—to provide these soldiers with sufficient knowledge of "American" ideals and the reasons for maintaining them. Consequently, it was concluded, such vulnerability can be reduced in the future if our schools and other institutions teach more explicitly the reasons for maintaining these beliefs. The present hypothesis predicts that, on the contrary, pre-exposure to the counterarguments themselves (in weakened form) is more effective as an immunizing agent than is the presentation of arguments supporting the cultural truism.

A second hypothesis follows from the postulate of defense-by-avoidance and its implication that the person, due to a lack of practice, is inept at defending his beliefs when the necessity arises. In the case of such beliefs, the requirement that the person actively participate, without guidance, in the defense should have a deleterious effect on the amount of immunity to later persuasion that the defense confers. Very likely the person cannot employ such an opportunity for participation effectively and the time could be more efficiently used in presenting him with already developed defensive material. Furthermore, such required participation should tend to be so difficult a task as to interfere with the reception of any defensive material presented, and to bring home to him forcibly how unable he is to defend his belief, thus tending to weaken rather than strengthen his confidence in it. Other theoretical formulations leading to this second hypothesis are considered in later discussion. Also to be considered later are some theoretical implications of the forced compliance studies and of studies involving active recitation during learning that might seem to lead to the opposite prediction.

A third hypothesis involves an interaction effect between the two main variables, type of defense and amount of participation. The requirement of active participation should be less detrimental with the supportive type of defense than with the type involving pre-exposure to counterarguments. This prediction was based on two considerations, one having to do with motivation, the other with the danger of overexposure to counterarguments, during the defense.

Although the second hypothesis points to a net negative effect from a requirement of active participation under the present situation, one component of the participation effect is likely to be beneficial, namely, its effect on motivation. By postulation, the person is not only unprepared, but also unmotivated to defend the supposedly unassailable cultural truisms prior to any attack. Hence, having to make an observable contribution during the defensive session should tend to supply some useful extrinsic motivation for working seriously at the defensive task. The lack of intrinsic motivation (and the consequent beneficial contribution of the requirement for active participation) is likely to be more pronounced in the supportive defense condition, which is likely to appear to involve belaboring the obvious. The pre-exposure to (weakened) counterarguments involved in the other defense suggests that the "truism" might indeed be assailable, and, hence, provides more intrinsic motivation for working seriously at the defensive task than does the supportive defense.

A second consideration leading to the predicted interaction effect involves the possibility that the active participation condition, being less under the experimenter's control, might overexpose the person to the dangerous material involved in the weakened counterarguments. This possible basis of the predicted interaction effect (as well as theoretical considerations leading to the prediction of an

opposite interaction effect) are discussed more fully in connection with the results obtained.

METHOD

General Procedure

The experiment was represented to the subjects as a study of the relationship between reading and writing skills. Each subject took part in two one-hour experimental sessions. The first was designed to give the subject various types of belief-immunizing treatments; the second, 2 days later, exposed him to strong counterarguing messages attacking the beliefs.

First session. In the first, or "immunizing," session, each subject was given defensive treatments on two beliefs. The first defensive treatment involved assigning him a belief that is almost universally held by his peers and instructing him (under experimentally varied conditions) to write a 20-minute essay defending that belief.

After the 20 minutes were up the subjects were given the second immunizing treatment, which involved reading for 5 minutes a 1,000-word essay (of an experimentally varied type) on another cultural truism and then answering a series of comprehension questions about the material read.

The subject's final task in the first session was to complete an attitude questionnaire designed to measure strength of belief in the cultural truisms. The subject was told to indicate his own belief, regardless of the positions taken in the essays that had been assigned to him, purportedly to allow us to check on whether the person's beliefs about the particular subject matter used in the test had any effect on his performance level. The belief scores obtained on the immunized issues in this first session questionnaire furnished a measure of the direct effect of the defensive treatments themselves, prior to the second session presentation of the strong counterargument messages.

Second session. The subject took part 48 hours later in a second session which involved presentation of three 1,000-word essays containing strong counterarguments against each of the two previously defended, and one additional belief. After being allowed 5 minutes for reading each of these three messages, the subject was required to answer a series of multiple-choice questions on their contents. Finally the subject was once again given the opinion questionnaire just described, and told to indicate his present personal belief, whatever might have been argued in any of the previous essays. The beliefs which the subjects indicated on these questionnaires on the immunized and unimmunized beliefs provided the measure of the effectiveness of the various immunizing conditions.

At the end of the experiment considerable effort was expended to inform the subject of the nature of the study and of the deceits used. Particular stress was put on the fact that the immunization and counterargument messages were both constructed for propaganda purposes and that no weight should be given to any argument simply because it had been included in one of these propaganda messages.

Independent Variables

Amount of active participation in the defense. The main manipulation of the degree of active, unguided participation was the already described reading vs. writing variation, the writing condition, of course, being defined as requiring more active participation than the reading. Within each of these conditions there was a further manipulation of the variable. There were passive and active subtypes of the reading condition: half of the subjects were told simply to read the material silently for 5 minutes, so as to be able later to answer questions testing their comprehension of it, the other half were instructed in addition to select and underline the one sentence in each paragraph most fully and succinctly summarizing that paragraph. The passages read were identical in active and passive subtypes, each consisting of a 1,000-word, five-paragraph essay defending one of the cultural truisms. The first paragraph summarized four arguments and each of the following paragraphs developed one of these in detail.[2] Within the writing condition also, there were two variations—guided and unguided writing. Half the subjects were given an outline summarizing four defensive arguments that could be used, and the other half received no such guiding outline.

Type of defense. Two types of defensive material were employed. In half the treatments, the prior defense involving presenting the subject with arguments supporting the cultural truism to which he initially adhered, in the other half, it involved exposure to possible counterarguments against the truism together with refutations of these counterarguments. In the reading conditions both of these types of defenses took the form of 1,000-word, five-paragraph essays, each essay presenting four supporting arguments, or mentioning and then refuting four counterarguments. In the writing-from-outline participation condition, the two types of defense involved presenting the subject either with one-sentence synopses of each of four supporting arguments or with four two-sentence synopses, each mentioning a counterargument and an argument refuting it. In the writing-without-outline participation condition, the subjects in the supportive conditions were instructed to write an essay giving arguments in support of the truism, and those in the refutational, an essay mentioning and refuting possible counterarguments against the truism.

Design

Each subject served in an experimental condition on each of four cultural truisms. On one, he received neither the defensive treatment in the first session nor the strong counterargument treatment in the second

[2] The complete text of all the communications, defensive and counterargumentative, used in this study has been deposited with the American Documentation Institute. Order Document No. 6745 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D.C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

(to provide an index of the base-levels of the beliefs). On another, he received no defensive treatment in the first session, only the strong counterarguments in the second (to provide a measure of the effect of the attack in the absence of prior defense). On the other two beliefs, he received both the prior defensive and subsequent strong counterargument treatment. The pair of defensive treatments given any one subject was not assigned completely at random: each subject received one reading and one writing defense. (Since 20 minutes were allowed for writing and only 5 for reading, this restriction was imposed to assure that all subjects served the same amount of time in the immunization session.) Hence, effects involving reading vs. writing conditions constitute intrasubject comparisons, while the other effects involve cross-subject comparisons. For this reason, in the results presented below, different parts of the variance are used as the "error" terms for the two classes of effects. The fact that this design required each subject to serve in four conditions is the reason that materials were developed on four different issues. These issues were rotated systematically around conditions from subject to subject.

Issues

On the basis of an earlier survey of student opinions (involving different subjects from those serving in the experiment) four issues were selected from an initial pool of 20, all of which dealt with health topics. Since the hypotheses dealt particularly with cultural truisms, the criteria for selection were extremeness and homogeneity of precommunication beliefs. The health area was found to be particularly fertile of such beliefs among college students. The beliefs selected were the following: "Everyone should get a chest X ray each year in order to detect any possible tuberculosis symptoms at an early stage"; "The effects of penicillin have been, almost without exception, of great benefit to mankind"; "Most forms of mental illness are not contagious"; "Everyone should brush his teeth after every meal if at all possible." For brevity's sake, these four issues are henceforth referred to by the following letters, respectively: T (tuberculosis), P (penicillin), M (mental illness), and D (dental).

The subjects indicated their beliefs in these statements by marking an "X" in the appropriate space on a 15-interval graphic scale that looked as follows:

Definitely FALSE	Probably FALSE	Uncertain	Probably TRUE	Definitely TRUE

The subjects were told to make an "X" in the space that best indicated their belief about the truth of the statement. The preliminary survey data showed that if we scale these intervals from 1–15 going towards the right, the mean belief on each of the four issues was over 13, with the mode at 15.

Subjects

A total of 130 subjects participated in both sessions of the experiment. These were all unpaid students in freshman rhetoric courses at the University of Illinois. Data for 16 additional subjects who took part in the first (immunization) session, but did not appear for the second (counterargument) session are omitted from the analysis. Since the rhetoric course is required of almost all freshman students who enter any college of the university, the subjects probably constitute a representative sample (as regards age, sex, intelligence, etc.) of first-semester freshman students in all colleges of this large state university.

RESULTS

General Effects

The control data from the first session indicate that the initial beliefs were extreme and homogeneous, amounting, as intended, to cultural truisms. Even though the beliefs were already at the high initial level of 13.26 points on the 15-point scale (see Table 1), the eight defensive conditions did further strengthen them, moving them up to about 50% of the remaining scale distance ($p < .001$).

Despite their unqualified initial acceptance the beliefs proved highly vulnerable to the massive counterarguments in the second session, which reduced them from the 13.26 point level of the first session to 6.64 points ($p < .001$) when not preceded by any immunization treatment. Even with prior immunization, the strong counterarguments lowered the beliefs considerably (see Table 2) though the mean effect of the eight immunization treatments was to attenuate their effect by about one-third ($p < .001$). These eight treatments differed considerably among themselves in regard to both direct strengthening effects (see Table 1) and immunizing effects (see Table 2). We turn now to this question of comparative effects.

Comparative Efficacy of Supportive and Refutational Defense

Both types of defense had significant direct strengthening effects on the beliefs, prior to the presentation of the strong counterargument messages but, as can be seen in Table 1, the supportive defense was slightly more effective ($p < .10$) in this regard than the refutational.

Regarding immunizing effectiveness, on the other hand, the refutational defense was superior ($p < .001$) to the supportive, as can be seen in Table 2. In fact, the supportive defense had hardly any immunizing effect. The 7.39 point level to which the strong counterarguments reduce the beliefs after the four supportive defenses is only slightly ($p = .16$) higher

TABLE 1
BELIEFS AFTER IMMUNIZATION BUT BEFORE EXPOSURE TO STRONG COUNTERARGUMENTS
(15 = absolute agreement; 1 = absolute disagreement)

Issues	Immunization Condition								Control (No Immunization)
	Reading				Writing				
	Supportive		Refutational		Supportive		Refutational		
	Passive	Underline	Passive	Underline	Outline	No outline	Outline	No outline	
D	14.75	15.00	15.00	14.50	14.11	13.83	13.57	14.57	13.39
M	15.00	14.22	13.14	12.71	14.62	14.33	13.00	12.22	13.22
P	14.67	12.62	14.67	15.00	14.11	14.11	13.71	12.67	13.41
T	14.22	14.89	14.22	14.14	14.83	14.38	14.30	14.50	13.02
Weighted mean	14.62	14.16	14.34	14.13	14.38	14.19	13.71	13.46	13.26
Total *N*	32	32	35	31	32	32	31	35	260

TABLE 2
BELIEFS AFTER BOTH IMMUNIZATOIN AND EXPOSURE TO THE STRONG COUNTERARGUMENTS
(15 = absolute agreement; 1 = absolute disagreement)

Issues	Strong Counterarguments Only (No Immunization)	Immunization Condition								Control (No Immunization or Counterarguments)
		Reading				Writing				
		Supportive		Refutational		Supportive		Refutational		
		Passive	Underline	Passive	Underline	Outline	No outline	Outline	No outline	
D	4.74	8.25	9.50	12.70	10.00	9.00	9.17	7.57	7.57	12.88
M	7.06	6.17	5.33	11.14	10.86	8.75	4.89	8.71	7.33	11.50
P	8.62	8.78	11.62	11.10	13.86	7.67	7.56	9.71	10.89	13.44
T	6.42	6.33	5.11	11.00	10.29	5.67	5.25	10.30	11.40	12.69
Weighted mean	6.64	7.47	7.63	11.51	11.13	7.94	6.53	9.19	9.46	12.62
Total *N*	130	32	32	35	31	32	32	31	35	130

than the 6.64 point level to which they were reduced after no defense at all. Even the most effective supportive defense (writing-with-outline) was superior to no-defense only on the .13 level of significance; and the poorest supportive defense (writing-without-outline) was actually (though trivially) worse than no defense at all, as can be seen in Table 2.

Comparative Efficacy of the Different Participation Conditions

The defensive treatments clearly lost effectiveness to the extent that they required the subject's unguided active participation in the defense. This detrimental effect is seen most clearly if we consider the major manipulation of participation, reading vs. writing. The reading conditions had a greater ($p = .05$) direct strengthening effect (see Table 1), and even more pronounced, a greater ($p < .001$) immunization effect (see Table 2).

All four of the comparisons between minor manipulations of participation were in the same direction. Passive reading proved superior to reading-and-underlining in regard to both direct strengthening and immunization. Likewise, writing-from-outline proved more effective than writing-without-outline in both these respects. While all four effects were in the same direction and agreed with the significant effect of the main (reading vs. writing) manipulations of participation their magni-

tudes (which can be seen in Tables 1 and 2) fall short of conventional significance levels.

Interaction between Type of Defense and Participation

The interaction between type of defense (supportive vs. refutational) and participation (reading vs. writing) is in the predicted direction but of a magnitude that attains only borderline significance. The superiority of reading over writing is greater with the refutational than with the supportive defense, in regard to both direct strengthening effect ($p = .10$) and immunizing effect ($p = .08$), as can be seen in Tables 1 and 2.

The interaction effect emerges more clearly when we contrast writing-from-outline—with regard to which there is a double reason for expecting the predicted interaction—with the other three participation conditions. This condition has the greatest immunizing effectiveness of the four with the supportive type of defense and the least of the four with the refutational defense, the interaction being significant at the .03 level.

DISCUSSION

In the case of each of the three predicted effects there were alternative theoretical bases for predicting the same effect or, in some cases, the opposite effect. These alternative formulations, together with subsidiary findings are considered in the following sections which take up each of the predicted effects in turn. The probable limits of valid generalization of each of the obtained effects are also discussed.

Type-of-Defense Effects

The finding that the direct strengthening effect was greater with presentation of supporting arguments than with pre-exposure to refuted counterarguments calls attention to the danger that the refutational defense (or any defensive procedure involving pre-exposure to the counterarguments, however weakened) might have the direct effect of slightly weakening the belief by acquainting the subject with unsuspected counterarguments. Should this be the case, a practical question would arise regarding the advisability of exposing a large population to such an immunization procedure when only a small, unpredictable sample is likely ever to be exposed to the subsequent massive dose of counterarguments. The data indicate, however, that no such direct "boomerang" effect occurred. On the contrary, the scores presented in Table 1 show that the refutational treatments had the immediate effect of strengthening the belief by 0.65 points ($p = .02$).

The finding of a reversed relationship between the immediate direct strengthening effect and the immunizing effectiveness of the refutational as compared with the supportive defense, raises the question of how well the immunizing effectiveness of a treatment can be predicted from its immediate apparent strengthening effect. The results over the eight defensive conditions indicate that immunization is very poorly predicted by the direct strengthening effect: over all eight defensive conditions, the rank order coefficient of correlation between immediate strengthening and subsequent resistance is actually negative ($\rho = -0.38$). Hence, the initial strength of belief or the amount of induced strengthening is a very poor indicator of conferred resistance to subsequent strong counterarguments. Some treatments (e.g., the refutational defenses) that appear to leave the beliefs relatively weak actually confer on them hidden reserves of resistance, while other treatments (e.g., the supportive defenses) that seem to have much strengthened the beliefs have actually left them "paper tigers."

The supportive vs. refutational manipulation in this study is quite similar to the one- vs. two-sided argument manipulation in previous studies (Hovland, Lumsdaine, & Sheffield, 1949; Lumsdaine & Janis, 1953). Their one-sided condition is much the same as our supportive; their two-sided condition is, however, a less close analog to our refutational. While their two-sided defense involved the supportive arguments plus some lesser mention of counterarguments either unsupported or refuted, our refutational defense involved no supportive arguments, only some mention of counterarguments with detailed refutations thereof. The obtained superiority of the refutational defense in this study suggests that the previously found superiority of the two-sided messages may have been due solely to their having pre-exposed the subjects to the weakened counterarguments.

without the concomitant exposure to the supporting arguments having been necessary. Whether the concomitant presentation of the supporting arguments does confer any additional effect, and under what circumstances—for example, before or after or intermingled with the weakened counterarguments—remains a question for subsequent research.

There are theoretical reasons for not generalizing the obtained immunizing superiority of the refutational over the supportive defense beyond certain of the conditions that obtained in the present experiment. One such restriction derives from the type of belief used, namely, those culturally regarded as being truisms almost beyond debate. Beliefs of this type are the most likely to have been overprotected against counterarguments in the past and, hence, particularly in need of resistance-stimulating pre-exposure to weakened forms of the counterarguments before massive exposure to strong forms. The moderation of less extreme beliefs may result from the obscurity or unimportance of the issue or, in the case of an important issue, to the person's having perceived reasons both for and against the belief. In the latter case, the person should be more prepared for strong counterarguments and, hence, the refutational defense may prove relatively less effective than we found with initially extreme beliefs.

A further restriction on the generality of the immunizing superiority of the refutational defense derives from the time parameters of the present study. Immunizing effectiveness was measured over the very brief interval of 48 hours. Presumably conferred immunity dissipates with time so that the effectiveness of all conditions would diminish as the interval between the defensive treatments and the presentation of the strong counterarguments lengthened. The question remains whether there might be different decay rates depending on how the immunity was conferred.

Another reservation about the obtained immunizing effectiveness of the refutational defense derives from the fact that in the present study the strong counterarguments subsequently presented to the subject were the very counterarguments previously refuted in the defensive session. The question arises whether the refutational defense has any general immunizing effect—lessening the subsequent belief-weakening impact not only of strong forms of the same counterarguments but of alternate counterarguments as well. Once again the medical field supplies an analogy in the question of whether inoculation with a weakened form of one strain of a virus produces immunity to other strains as well. At least two possible bases for expecting a generalized immunity effect from exposure to refutational arguments can be suggested. First, pre-exposure to the weakened counterargument may, by making the subject more aware of the vulnerability of his belief, stimulate him to develop supporting arguments and to think up and refute other counterarguments. Second, the earlier experience of seeing some counterarguments refuted may lessen the impressiveness of all subsequently-received counterarguments against the belief.

The refutational defense used in the present experiment constitutes just one example of a larger class, namely, immunization by pre-exposure to the counterarguments in weakened forms, thus stimulating, without overwhelming, the subject's defenses before he is exposed to strong forms of the counterarguments. The counterarguments might be weakened during the inoculation treatment by less drastic procedures than the explicit refutations used here. They might, for example, be presented as emanating from a negatively valenced source or as endorsed by a negative reference group. Or opposition to them might be registered as coming from the subject's peer groups, or prestige figures, or the general consensus. Alternatively, these counterarguments might simply be stated without supporting evidence or with patently specious supporting evidence. Use of these alternative treatments involving pre-exposure to arguments that are weakened less completely than the explicit refutation carries the risk that the counterarguments might be so strongly presented in the immunizing session that they actually weaken the belief. On the other hand, the exposure to somewhat stronger forms than used in this experiment might be more stimulating to the person's defenses, and thus develop greater resistance to a subsequent massive dose of counterarguments.

This question, concerning the optimal

strength of exposure to counterarguments during the immunizing session, raises a theoretical issue regarding the underlying mechanism of the immunity conferred. (The same issue also arose in the discussion of the specificity of the conferred immunity—whether just against the same counterarguments to which pre-exposed, or others as well.) One possible mechanism is that pre-exposure strengthens the belief by stimulating the subject to develop supporting arguments for maintaining his belief. Alternatively, the prior experience of seeing the weakness of counterarguments may lessen the perceived credibility of the later strong forms of counterarguments. If the former mechanism is paramount, pre-exposure to relatively strong counterarguments (up to the point of strength where they begin to overwhelm, rather than stimulate the subject's defenses) would be optimal. If the latter mechanism is the more important, then the more weakened the counterarguments during the pre-exposure, the greater the conferred immunity.

Effects of Requiring Participation

The superiority of the low over high participation conditions with respect to both direct strengthening and conferred immunity was found in all six of the possible independent comparisons. Since the direction of this effect is so clear-cut and yet might seem unexpected in the light of some previous studies on the effects of active participation, discussion seems in order regarding the postulated mechanisms and regarding relevant distinctive features of the present study.

The hypothesis that the defense of a cultural truism would be less effective to the extent that the believer is required to participate in it actively was based on the assumption that the believers are poorly prepared for such a task. The assumption receives support not only from the confirmation of the hypothesis, but also from the incidental evidence that the obtained participation was very poor in both quantity and quality. In the no outline condition, the essays were typically very meager. About half of the subjects (34 out of the 67) in this condition produced only one or no defensive argument. In the outline condition, the arguments tended to be unelaborated reiterations of the points presented in the outline. The poor quality of these essays leaves a strong impression that the defensive session tended to be largely wasted on the writing condition. Indeed, some suggestion is offered that the poorness of the subject's performance may have raised doubts in his own mind regarding the validity of his belief, in the finding (see Table 2) that after the subject was called upon without the benefit of any guiding outline to write a defensive essay giving arguments in support of his belief, the strong counterarguments reduced the beliefs more than after no defense at all.

A related basis for the obtained detrimental effect of requiring active participation derives from the limited time allowed for the defense in this study. Given that the person was poorly prepared to exercise initiative in defending his overprotected beliefs, the mechanics of the participation task may have usurped time needed for elicitation and absorption of the defensive material itself. Such stimulus-reception interference due to the requirement of participation could account for the superiority of passive reading over active (which required the subject to analyze and select the crucial clause in each paragraph, besides reading the material within an only moderately long, 5-minute interval). This factor could also be involved in the obtained superiority of both reading conditions over the writing conditions. Despite the quadruple time allowance of 20 minutes in the writing condition, it was apparently felt by the subjects to be less adequate for the task than was the 5-minute period allowed for the reading task. On a post-experimental questionnaire, 46 of the 130 subjects complained about insufficient time for writing the essay, and only 18 for the reading task ($p < .01$, by a chi square test). However, while this situation raises the theoretical question of whether the writing condition might prove relatively more effective should the time restriction be removed, it leaves a clear answer of practical importance: even though the time allowance was four times as great in the writing as in the reading conditions, the writing conditions still produced less immunity than the reading.

Another, alternative interpretation of the obtained superiority of the reading over writing

conditions has been suggested by Carl Hovland (in a personal communication). Hovland suggests that this manipulation may have affected perceived source credibility and, hence, (Hovland & Weiss, 1951), the persuasive impact of the defensive messages. Even though no explicit indication of message source was given in any of the immunization communications, there was probably a strong tendency for the subject to assume that the essays presented for reading came from a high prestige source since their subject matter was rather technical and involved specialized information, and, stylistically, they were well organized and literate. Since in the writing conditions, the arguments came from the subject himself (who was likely to be well aware of his lack of competence in the area) they were unlikely to gain as much from source credibility as in the reading condition. Even in the writing-from-outline condition in which the arguments were presented to the subject, they were presented in a telegraphic style that quite likely detracted from their prestigefulness. The factor of source prestige may at least partially explain the obtained superiority of the reading over writing conditions. However, this explanation accounts less adequately than does the earlier one for the findings in forced compliance experiments (discussed below) of the opposite effect. In those studies active participation was found to augment persuasive effectiveness even though the experiments do not differ greatly from the present one in regard to source-credibility.

In generalizing the obtained negative relation between immunizing effectiveness and amount of active participation in the defense, it is important to note that the present study involved beliefs with which the subject was in complete agreement at the outset and which were practically unquestioned in his culture so that the subject was unpracticed in their defense. There are two lines of research that lead us to expect that the relative value of active participation would be greater in the case of controversial, frequently argued beliefs of lower initial strength.

A number of studies have shown that in situations of forced compliance, active participation is more efficacious than passive reading in converting the person to a new belief that had initially been rejected. Kelman's work (1953) on response restriction, Janis and King's (1954) and King and Janis' (1956) work on active role playing and improvisation, Festinger's (1957) work on behavioral commitment and Lewin's (1958), and Hovland, Campbell, and Brock's (1957) work on public decision all furnish instances of the indoctrination value of overt participation. Likewise, practitioners of consent engineering, including essay contest sponsors, evangelists, indoctrinators of prisoners-of-war, etc. have long been acting on the postulate that overt participation increases opinion change. In these previous instances, however, the defended belief is rather disparate from the subject's own initial belief whereas in the case of the cultural truisms used in this experiment the subject tends to be in such completé initial agreement with the proposition that little potential remains for further internalization from the active participation. Also, in the studies cited the active participation requirement usually came just after the subject received a persuasive message that served as a guide for his participation.

A second line of research, from the learning area, also suggests that it would be unwise to generalize the present finding of a detrimental effect from active participation beyond the case of infrequently argued beliefs. The classic study by Gates (1917) and several subsequent studies have demonstrated that a greater amount of learning occurs under conditions of active recitation than of passive reading. Since opinion change is positively related to message comprehension (McGuire, 1957), the Gates study suggests that active participation in the defense would produce greater strengthening of the belief. That the opposite result was found is not surprising in view of the type of material involved in the present participation task. Since the subjects were unpracticed in defending the beliefs, the task was difficult and they were on an early segment of the learning curve, conditions under which the learning contribution of active participation has been found to be small or even negative (Kimble & Wulff, 1951; Krueger, 1930). Furthermore, the material was meaningful, which again has been found to diminish the benefits from active participation (Peterson, 1944).

Interaction Effect

The interaction hypothesis received the least impressive confirmation of the three predictions. However it received its most substantial support when the writing-from-outline participation condition, in which both the postulated mechanisms (lack of need for extrinsic motivation and danger of overexposure to the counterarguments) are operative, is contrasted with the other three participation conditions. This finding suggests that both mechanisms contribute to the effect, since the condition in which only one mechanism (lack of extrinsic motivation) is operative, writing-without-outline, falls between this condition and the conditions (reading) in which neither mechanism is operative, in regard to the superiority of refutational over supportive defense.

The interaction effect predicted and obtained in this study cannot be generalized to situations that fail to present the refutational material in such wise that the subject perceives it as a defense of his belief, and not an attack on it. (There is always some danger, which must be minimized for the present interaction effect to occur, that the refutational defense, since it mentions the counterarguments at the outset, may be misperceived by the subject as an attack on his belief.) The basic postulate of defense-by-avoidance underlying the present study would indicate that where the refutational defense is likely to be initially misinterpreted by the subject as an attack on his belief, the subject should tend to avoid exposing himself to it (Janis, 1957). Hence, in such ambiguous situations the opposite direction of interaction would be predicted, namely, that the extrinsic pressure to expose oneself to the defensive material, in the requirement of overt participation, would be more important with the refutational than the supportive defense. Such interaction should be increasingly pronounced as the subject's confidence in his belief declines (Brodbeck, 1956; Festinger, 1957).

SUMMARY

It was postulated that people characteristically defend their beliefs (particularly cultural truisms) by avoiding exposure to contradictory information, with the result that in forced exposure situations these untried beliefs prove highly vulnerable to persuasion. Since the person is unmotivated to develop a defense of his belief to the extent that he considers it invulnerable, it was hypothesized that such beliefs are more effectively immunized against persuasion by pre-exposure to counterarguments (in a weakened form that stimulates, without overcoming, the receiver's defenses) than by pre-exposure to arguments supporting the belief. Second, since the person is unpracticed in the defense of such beliefs, it was predicted that the immunizing pretreatments would lose effectiveness to the extent that they required the person to participate actively, without guidance, in the defense. Finally an interaction effect was predicted such that the detrimental effect of requiring active participation is greater with the defense involving pre-exposure to the counterarguments than with the supportive defense.

The subjects were 130 college freshmen who took part in two experimental sessions. In the first, they received immunizing treatments on various strongly held health beliefs. Type of defense (pre-exposure vs. supportive) and amount of participation in the defense were varied orthogonally. The subjects' beliefs were then measured to ascertain the direct effects of the defensive treatments. In the second session, the subjects were exposed to strong counterarguments on both immunized and nonimmunized beliefs and then the final beliefs were measured to ascertain the comparative immunizing effectiveness of the defensive treatments.

Both of the predicted main order effects were confirmed at the .001 level. Some support ($p = .08$) was also found for the predicted interaction effect. The defensive treatments, besides conferring immunity against subsequent counterarguments, also had an immediate direct strengthening effect on the beliefs ($p = .01$), but the amount of direct strengthening was not predictive of the latent resistance to counterarguments conferred by these treatments.

REFERENCES

BERELSON, B., LAZARSFELD, P. F., & MCPHEE, W. N. *Voting*. Univer. Chicago Press, 1954.

BRODBECK, MAY. The role of small groups in mediating

the effects of propaganda. *J. abnorm. soc. Psychol.*, 1956, **52**, 166–170.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

GATES, I. A. Recitation as a factor in memorizing. *Arch. Psychol., NY*, 1917, **6**, No. 40.

HOVLAND, C. I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *Amer. Psychologist*, 1959, **14**, 8–17.

HOVLAND, C. I., CAMPBELL, ENID H., & BROCK, T. Effect of "commitment" on opinion change following communication. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* Vol. 1. New Haven: Yale Univer. Press, 1957. Pp. 23–32.

HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments on mass communication.* Princeton: Princeton Univer. Press, 1949.

HOVLAND, C. I., & WEISS, W. The influence of source credibility on communication effectiveness. *Publ. opin. Quart.*, 1951, **15**, 635–650.

JANIS, I. L. Motivational effects of different sequential arrangements of conflicting arguments: A theoretical analysis. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* Vol. 1. New Haven: Yale Univer. Press, 1957. Pp. 170–186.

JANIS, I. L., & KING, B. The influence of role-playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211–218.

KELMAN, H. C. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, **6**, 185–214.

KIMBLE, G. A., & WULFF, J. J. *Participation and guidance during the learning session.* (Audio-vis. Div., HRRL, AFRDC) Washington, D. C.: Bolling Air Force Base, 1951.

KING, B., & JANIS, I. L. Comparison of the effectiveness of improvised versus non-improvised role-playing in producing opinion change. *Hum. Relat.*, 1956, **9**, 177–186.

KINKEAD, E. *In every war but one.* New York: Norton 1959.

KLAPPER, J. What do we know about the effects of mass communication: The brink of hope. *Publ. opin. Quart.*, 1957, **21**, 453–474.

KRUEGER, L. C. The relative effect of interspersing a recall at different stages of learning. *Arch. Psychol., NY*, 1930, **18**, 15–25.

LEWIN, K. Group decision and social change. In Eleanor Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* New York: Holt, 1958. Pp. 197–211.

LUMSDAINE, A. A., & JANIS, I. L. Resistance to "counter-propaganda" produced by one-sided and two-sided "propaganda" presentation. *Publ. opin. Quart.*, 1953, **17**, 311–318.

MCGUIRE, W. J. Order of presentation as a factor in "conditioning" persuasiveness. In C. I. Hovland (Ed.), *The order of presentation in persuasion.* Vol. 1. New Haven: Yale Univer. Press, 1957. Pp. 98–114.

PETERSON, H. A. Recitation or recall as a factor in the learning of long prose selections. *J. educ. Psychol.*, 1944, **35**, 220–228.

UNITED KINGDOM MINISTRY OF DEFENSE. *Treatment of British prisoners of war in Korea.* London: HM Stationery Office, 1955.

UNITED STATES SENATE (84th Congress, 2nd Session), Committee on Government Operations, Permanent Subcommittee on Investigations. *Communist interrogation, indoctrination and exploitation of American military and political prisoners.* Washington, D. C.: United States Government Printing Office, 1956.

(Received February 23, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 338-345

AFFECTIVE TONE DURING ENVIRONMENTAL REGULARITY AND CHANGE[1]

SALVATORE R. MADDI[2]
University of Chicago

A number of writers interested in understanding abnormality have clearly implied that unexpected occurrences are anxiety-provoking (Fenichel, 1945; Kelly, 1955; Sullivan, 1954). That this is too extreme a position is suggested by Hebb's (1955) discussion of the pleasantness of mild risk-taking and gambling. McClelland, Atkinson, Clark, and Lowell (1953) have incorporated both possibilities in their hypothesis that a small discrepancy between expectancy and occurrence arouses positive affect leading to approach behavior, while a large discrepancy arouses negative affect leading to avoidance behavior. Their emphasis is on the degree of difference between what is expected and what actually occurs, or on the unexpectedness of the occurrence.

McClelland et al. (1953) have interpreted a number of findings at the animal as well as at the human level in terms of their hypothesis, and George and Bonney (1956) found it useful in explaining results obtained by Siipola (1950) in an investigation of emotional reactions to color on the Rorschach. Bindra and Spinner's results (Bindra, 1959) also may be interpreted as consistent with this position. They found that exposure to a greatly changed form of a familiar stimulus situation produced grooming in rats, while exposure to an only moderately changed form produced sniffing and locomotion. While the latter behaviors resemble approach in that they are directed toward the stimulus situation, the former is perhaps more reasonably considered avoidance as it is not directed toward the situation. Presumably, the greatly changed situation was more unexpected than was the less greatly changed one. That the explanatory value of McClelland's position has not been unequivocally established is indicated by results such as those reported by Kenny (1955). His study of appreciation of joke endings as a function of their unexpectedness is only partially consistent with this position.

McClelland et al. (1953) also state that when expectancy and occurrence do not differ, there is no affective arousal and no behavior that can be characterized as approach or avoidance, unless the behavior is caused by some other affect-arousing property of the situation. However, completely expected situations at least sometimes seem to lead to boredom, anger, or avoidance. For example, Karsten (1928) reported that the effects of continued repetition of the same activity include an increase in emotional outbursts, the tendency to vary the task and concentrate on something else, and the attractiveness of other activities. It is proposed that both greatly unexpected and completely expected situations arouse negative affect, while moderately unexpected situations arouse positive affect. This statement would apply most clearly in situations in which rewards and punishments, as usually conceptualized, are at a minimum.

This paper reports the test of some hypotheses following from the proposition just stated. A number of definitions and assumptions must be considered before proceeding to the experiment. As the expectancy is based on temporal or sequential relationships among events, the more precise term *series of events* will be employed instead of *stimulus situation*. A series of events is a group of stimuli that form, or are arranged in, an order according to some temporal principle. A repetitive series consists of events that are similar to, and have the same sequence as, those of a previous series. The occurrence of two or more successive series that do not differ constitutes *environmental regularity*. A nonrepetitive series is one that differs from the previous series in order of events, content, or both. The occurrence of a nonrepetitive series following en-

[1] The research reported here is from the author's (1959) doctoral dissertation done at Harvard University. The preparation of this paper was facilitated by funds from Ford Foundation Behavioral Sciences Grant #1-5600-49-2863 (Maddi).

[2] The author wishes to express his gratitude to David C. McClelland and Richard L. Solomon for their aid and encouragement, to Donald W. Fiske for critical reading of this paper, and to Lindsay Churchill and John Shaffer for their assistance.

vironmental regularity is called *environmental change*. The *degree of change* constituted by a nonrepetitive series refers to the difference between it and the repetitive series it follows. An expectancy is the cognitive representation of a series of events, which permits prediction of the order in which the events occur.

In the experiment reported below, an initially novel series of events was repeated a number of times (environmental regularity). As a series is repeated, its unexpectedness should decrease. The first hypothesis to be tested is that as a series is repeated, affective tone is initially negative, becomes positive, and then becomes negative again. After a number of series repetitions, a final series of events was presented. This final series was repetitive (no change) for some subjects, and nonrepetitive (environmental change) for other subjects. Of the latter subjects, some experienced a relatively large degree of change, while others experienced a relatively small degree of change. The larger the degree of change present in the final series, the more unexpected it should be. The second hypothesis to be tested is that conditions of no change, and of relatively large degree of change are associated with negative affect, while a relatively small degree of change is associated with positive affect.

METHOD

Groups and Instructions

Each subject, a paid volunteer from a population of college students, was randomly assigned to one of three groups. Each group consisted of 48 subjects. Four subjects from the same group were tested during each testing session. Prior to the beginning of the experiment, each subject was provided with a booklet, of which the first page and the last two pages were blank. On every remaining page there was either a number or a sentence completion stem. The subjects were told that they were participating in a sentence completion task, and that they were to turn a page of the booklet each time the signal was given. Whenever they encountered a stem they were to complete the sentence in any manner they wished. They were also told that numbers would appear, and that these had been included as spacers between the stems so that they might rest when not completing sentences.

While two groups received only these instructions, the third group was also asked to predict the occurrence of types of events. They did this by writing an "N" (number) or an "S" (stem) on the bottom of each page depending upon what type of event they thought would occur next. As will be described below, this group experienced environmental regularity. It was called the Prediction group. One of the other two groups that did not predict also experienced regularity. It was called the No Prediction group. The remaining group did not predict and did not experience regularity. Included for control purposes, it was called the No Regularity group.

Environmental Regularity

The experimental treatment described in this, and the following section is summarized in Table 1. As they turned the pages of their booklets, all subjects encountered nine stems. In all booklets, the first stem was preceded by at least one number. The unit formed by the first stem and the numbers antecedent to it was considered to be the first series of events. Each of the next seven stems was separated from the preceding stem by at least one number. The units formed by each of these stems and the numbers separating it from the preceding stem were also considered to be series of events.

The first series experienced by the subjects in the Prediction and No Prediction groups consisted of a number, a number, and a stem, in that order. This sequence was repeated in the next seven series. Both groups experienced environmental regularity during this part of the experiment.

No two consecutive series experienced by subjects of the No Regularity group had the same sequence of events. Hence, this group did not experience environmental regularity. For each subject, the number of events (numbers) antecedent to the stem in each of the first eight series was chosen at random from between one and six, with the restrictions that no series with the same number of antecedent events should appear more than three times in the eight series, and that the number of antecedent events throughout the first eight series should equal the same sum for subjects of the Prediction and No Prediction groups.

TABLE 1
SUMMARY OF EXPERIMENTAL TREATMENT

Series	Group		
	Prediction	No Prediction	No Regularity[a]
1	NNS[b]	NNS	NS
2	NNS	NNS	NNNS
3	NNS	NNS	NNS
4	NNS	NNS	NNNS
5	NNS	NNS	NNS
6	NNS	NNS	NS
7	NNS	NNS	NNNS
8	NNS	NNS	NS
	Subgroups of Each Group[c]		
	A	B	C	D
9	NNS	NS *or* NNNS	S *or* NNNNS	NNS

[a] The number of events antecedent to the stem in each of the first eight series was determined randomly for each subject in this group. An example of one such random pattern is given.

[b] "N" represents a number, while "S" represents a sentence completion stem.

[c] Each group was subdivided into four subgroups. Each of the subgroups of a group received a different version of the final series.

Environmental Change

Following the first eight series, all subjects received a final series. Each group was subdivided into Subgroups A (8 subjects), B (16 subjects), C (16 subjects), and D (8 subjects). Each subgroup in a group received a different version of the final series (treatment). Treatment O was a series consisting of a number, a number, and a stem, in that order. Subgroup A of each group received this treatment. It constituted no change (a continuation of regularity) for the subjects of the Prediction and No Prediction groups, because its events had the same order as, and were similar to, those of previous series. Treatment 1 consisted of either a number followed by a stem, or of three numbers and a stem in succession (i.e., number, number, number, stem). In each group, one-half of Subgroup B received the former series, while the other half received the latter series. Treatment 1 constituted one degree of change for the subjects of the Prediction and No Prediction groups, because the stem was displaced either one position forward or backward from its usual position. Treatment 2 consisted of either a stem alone, or of four numbers and a stem in succession. In each group, one-half of Subgroup C received the former series, while the other half received the latter series. This treatment constituted two degrees of change for the subjects of the Prediction and the No Prediction groups, because the stem was displaced either two positions forward or backward from its usual position.

In the Prediction and No Prediction groups, Subgroups A, B, and C experienced, respectively, 0, 1, and 2 degrees of change in order of events. Although the comparable subgroups of the No Regularity group received the same treatments during the final series, they did not experience environmental change as it has been defined because they had not previously experienced regularity.

Treatment 3 consisted of two numbers and a stem in succession, but the stem was printed in much larger type than had been the case previously. Included for exploratory purposes, this treatment was experienced by Subgroup D of each group. While the subjects of the Prediction and No Prediction group who received this treatment experienced no change in order of events, there was a change in size of one of the events.

Stimuli

The specific numbers presented to each subject were selected at random from those between 10 and 57, with the restriction that no subject should experience any number more than once. The stems utilized in the first eight series were "In winter," "The stranger," "When he heard," "I feel," "Americans," "He thought," "My memories," and "When John met." The order of presentation of the stems was determined by an 8 × 8 latin square with the property that each stem was followed by each other stem only once throughout the eight possible orders. Each subject in a group was assigned one of the eight orders in such a manner that every subject experiencing a particular final series had received a different order. The stem "Bill seemed" was utilized in the final series for all subjects.

During the experiment, the experimenter sat behind the subjects and timed the exposure of stimuli with a stopwatch. The subjects of the No Prediction and No Regularity groups were given the signal to turn the page after 30 seconds in the case of stems, and after 10 seconds in the case of numbers. The subjects of the Prediction group were given an additional 5 seconds per stimulus in which to predict.

Minimization of Rewards and Punishments

As discussed earlier, the effects of the unexpectedness of the series upon affective tone should be most apparent when the other affect-arousing properties are at a minimum. Potential rewards were minimized in the following manner: the experimenter did not converse with the subjects, nor did he mention such matters as the importance of either their participation or of the experiment itself. The $1.00 each subject received, while it was a reward, cannot be considered to have had differential effects upon the various aspects of the subject's performance. Potential punishments were minimized by ensuring the subject's anonymity, by putting his participation essentially on a voluntary basis, and by testing him as part of a group. In addition, the instructions were not particularly ego involving, and an attempt was made to select nonthreatening stems. Nonetheless, one source of reward and punishment was the possible significance of a particular stem for a particular subject. The control of this factor will be discussed later.

Measure of Unexpectedness

It has been assumed that a series becomes less unexpected as it is repeated, and that a nonrepetitive series is unexpected to the degree that it differs from the repetitive series it follows. In this experiment, inaccuracy of prediction can therefore be considered an index of unexpectedness. Such a measure can be obtained for the prediction group as a whole by computing the percentage of inaccuracy in predicting the events comprising a particular series. For any series, this percentage is based on the proportion of inaccurate predictions to the total number of predictions made by the group.

No such measure was available for the No Predictions and No Regularity groups. However, a very indirect check on the assumptions was available. Following the experiment, all subjects were asked to indicate whether they had found it possible to anticipate when there would be a stem on the next page, and to state, if possible, how many numbers preceded the stem in each series but the last.

Measure of Affective Tone

The sentence completions were used as indicants of affective tone. This usage appears reasonable, as Bousfield (1950) has shown that pleasant and unpleasant moods determine the affective value of word associations. It was assumed that the affect evidenced in a given sentence completion would be a function of the significance of the particular stem for that subject and the degree of unexpectedness associated with the series of which the stem was a part. Investigation of the second factor alone was permitted by the counterbalancing of stem position, the assignment of stem orders to subjects at random, and the analysis of data by group and subgroup only.

Affective tone was considered to vary along a five-point scale, with 1, 2, 3, 4, and 5 representing, respectively, strongly negative affect, slightly negative affect, neutral affect, slightly positive affect, and strongly positive affect. On the basis of pilot study data, four such bipolar, five-point dimensions of affectively-determined content—to be described—were isolated. Each sentence completion was scored for affect on the dimensions relevant to it, and the resulting scores were combined into a composite affect score for that completion. These composite affect scores were used in testing hypotheses.

The composite affect score was assigned as follows: A completion to which none of the dimensions was relevant received a composite affect score of 3. If only one dimension applied to a completion, the completion was given the affect score obtained on that dimension. However, a completion could occasionally be scored toward both the positive and negative poles of the one dimension applying to it. If the intensity involved in both scores was the same (i.e., 1 and 5, or 2 and 4), the completion was given the composite affect score of 3. If the positive and negative scores did not agree in intensity (e.g., 1 and 4), the composite affect score was determined in the direction of the stronger of the two (in the example given, the final score would be 2). Occasionally, two dimensions were relevant to a particular completion. These cases were treated, in effect, as if the completion had been scored toward both the positive and negative pole of one dimension.

The dimension of *expected outcome* (negative pole is pessimism, and positive pole is optimism) was relevant whenever some future event or condition was referred to, or definitely implied in the completion, regardless of the unit of time involved. Completions scored on this dimension include "*I feel* that the world will end in destruction" (affect score of 1) and "*Americans* will triumph over the forces of Russia" (affect score of 5). The dimension of *evaluation* (unfavorable evaluation-favorable evaluation) was relevant whenever the completion indicated a point of view concerning a situation, event, or person. Completions scored on this dimension include, "*He thought* the experiment was nonsense" (affect score of 1), and "*Bill seemed* to be in favor of American foreign policy" (affect score of 5). The dimension of *interest* (indifference-enthusiasm) was relevant whenever there was some indication of the subject's involvement in the task, or in the subject matter of the completion. Completions scored on this dimension include, "*I feel* bored" (affect score of 1), and "*In winter* I am very active in work and play" (affect score of 5). The dimension of *description* (agreeable description-disagreeable description) was relevant whenever the completion described persons, conditions, or events, whether real or hypothetical. Completions scored on this dimension include "*The stranger* had a big knife" (affect score of 1), and "*In winter* the snow swirls through the streets covering everything with a blanket of lovely whiteness" (affect score of 5). The dimension of *expressed feeling* (unpleasant feeling-pleasant feeling) was relevant whenever the subject referred to his own feelings, or to those of others. Completions scored on this dimension include, "*He thought* the anxious, sinking feeling would never leave" (affect score of 1), and "*I feel* happy" (affect score of 5). In keeping with usual interpretation, expressive disturbances, such as failure to complete the stem, erasures, and cross-outs, were considered indications of strongly unpleasant feelings.[3]

The composite affect scores attributed to completions by the experimenter were used to test hypotheses. While scoring a particular completion, the experimenter did not know the group or subgroup to which it belonged, or (with the exception of completions involving the stem "Bill seemed," which was only used during the final series) during which of the nine series it had been written. In order to check reliability, the completions were scored by another person who, although trained by the experimenter, had no knowledge of the design of the experiment or of the hypotheses being tested.

RESULTS

Interscorer Reliability

The product-moment correlation between the composite affect scores attributed to the separate sentence completions by each of the scorers was .91, indicating that the scoring system is sufficiently objective to be employed effectively.

Environmental Regularity

The percentage of inaccurate predictions made by the prediction group in each of the first eight series in succession was 76, 47, 33, 16, 16, 10, 10, and 6. The decrement in inaccuracy that occurred during this period of regularity is significant at the .001 level ($\chi^2 = 62.88$; $df = 7$). These results indicate that in this group the unexpectedness of the first series decreased as it was repeated.

Comparable data are not available for the other two groups. However, an indirect check on the assumption that regularity reduces unexpectedness can be made. Forty-one of 48 subjects in the prediction group correctly answered the question concerning how many numbers had immediately preceded each stem but the last ($\chi^2 = 22.68$; $p < .001$). In the No Prediction group, 37 of 48 subjects answered correctly ($\chi^2 = 13.02$; $p < .001$). These results suggest that the conditions of the experiment were sufficient to establish an expectancy for both of these groups. This question is not an adequate basis for comparison of the groups experiencing regularity and the No Regularity group, because the subjects of the latter group could not answer correctly by stating any single number. But the question concerning the ability to anticipate stems does provide

[3] For a more complete description of the scoring system, including examples, see Maddi (1959).

some basis for comparison. In the No Regularity group, 8 of 48 subjects felt that they could anticipate stems. In contrast, 42 of 48 subjects in the Prediction group, and 38 of 48 subjects in the No Prediction group felt that they could anticipate stems. The corresponding 2×2 contingency table yields a chi square value of 62.33 ($p < .001$), suggesting that conditions of the experiment were not sufficient to establish an expectancy in the No Regularity group. Taken together, these results suggest that in both the Prediction and No Prediction groups environmental regularity reduced the unexpectedness of the series that was repeated.

An assumption implicit in the design of the experiment is that the period of regularity lasted long enough to reduce unexpectedness to a near-minimal value. As the inaccuracy of prediction on the eighth series was 6%, this assumption seems warranted for the Prediction group. There are no comparable data for the No Prediction group. However, there was only a small difference between these two groups in the number of subjects who could correctly answer the question concerning how many numbers had immediately preceded each stem but the last. The 2×2 contingency table comparing the Prediction and No Prediction groups with respect to correctness of prediction yields a chi square value of 0.62 ($p > .75$), suggesting that the assumption is warranted for the No Prediction group as well.

Figure 1 shows the relationship between mean affect scores and number of series experienced. The affect curves provide a test of the first hypothesis, which states that as the series is repeated, affective tone is initially negative, becomes positive, and then becomes negative again. The curves of both groups experiencing regularity are consistent with this hypothesis, while the curve of the No Regularity group is not.

The procedure followed in determining whether the curvilinearity shown in the data of the groups experiencing regularity could have occurred by chance was essentially that of comparing the two highest points on each curve with the two points on either side of them.[4] For the Prediction group, the sum of each subject's affect scores on the third and

[4] The author wishes to thank Frederick Mosteller for suggesting this procedure.

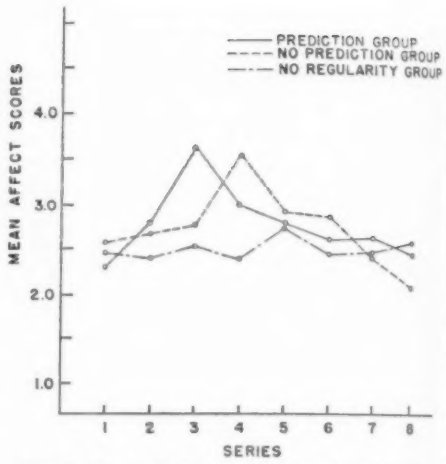

FIG. 1. Mean affect scores as a function of number of series experienced. (These series constituted environmental regularity for the Prediction and No Prediction groups.)

fourth series (called his Sum B) was compared with the sum of his scores during the first and second series (called his Sum A), and with the sum of his scores on the fifth and sixth series (called his Sum C). Each time Sum B was larger than were both Sum A and Sum C, the subject was given a comparison score of 2. Whenever Sum B was greater than one, but not both of the other sums, the subject was given a comparison score of 1, and whenever Sum B was the smallest, a score of 0 was recorded. If Sum B was tied for largest with one other sum, a score of 1.5 was given, and if Sum B was tied for smallest, a score of 0.5 was recorded. The same procedure was followed for subjects of the No Prediction group, except that Sum B was computed from the individual's affect scores during the fourth and fifth series. Accordingly, Sum A involved the second and third series, and Sum C the sixth and seventh series.

The mean comparison score that would be expected by chance is 1.00, since an individual's affect curve could take any shape, and, hence, he was as likely to obtain one of the possible comparison scores as another. For the Prediction group, the mean and standard deviation of comparison scores were, respectively, 1.28 and 0.68. The No Prediction group mean was 1.26, with a standard deviation of 0.86. The mean is significantly greater than

chance expectancy in the Prediction group ($t = 2.80$; $df = 40$; $p < .01$) and the No Prediction group ($t = 2.17$; $df = 40$; $p < .05$). These results indicate that in both groups, the curvilinearity could not very likely have occurred by chance alone.

The peak of the Prediction group's affect curve occurred in the third series, during which the group was inaccurate in 33% of its predictions. While inaccuracy of prediction decreased to 6% during the next five series, affective tone went from positive to negative. While inaccuracy of prediction decreased from 76% to 33% during the first three series, affective tone went from negative to positive. The affect curve of the group can be understood in terms of the proposed effects of decreasing levels of series unexpectedness upon affective tone.

Environmental Change

Treatments 0, 1, and 2 in the ninth series constituted 0, 1, and 2 degrees of change in order of events for the Prediction group. The percentages of inaccurate predictions made during Treatments 0, 1, and 2, were, 8, 42, and 64, respectively. This trend, significant at the .001 level ($\chi^2 = 16.40$; $df = 2$), is consistent with the assumption that the series is more unexpected the larger the degree of change it constitutes.

The most direct test of the second hypothesis involves the mean affect scores associated with Treatments 0, 1, and 2 in the final series. But in such an analysis each mean would be determined from the data of different subgroups which had not been equated for level of affect prior to the final series. A more accurate estimate of the effects of the various final series treatments involves partialing out the effects of prior level of affect. This was done by subtracting each subject's affect score on the eighth series from his score on the final series. When the result was a positive number, affective change toward the positive pole of the affective tone dimension had occurred; when the result was a negative number, negative affective change was inferred.

In order to establish the relevance of affective change scores, some implications of the second hypothesis will be made explicit. A precondition for the applicability of this hypothesis is that the period of regularity has lasted long enough to reduce unexpectedness to a near-minimal value. Following such a period of regularity, affect should be in the negative range. The occurrence of another repetitive series (i.e., no change) should produce negative change in affect. The occurrence of a moderate degree of change should produce positive affective change, while a large degree of change should produce negative affective change.

Figure 2 shows the relationship between mean difference scores and Treatments 0, 1, and 2, which represent, respectively, 0, 1, and 2 degrees of change in order of events for the Prediction and No Prediction groups. The difference between the means associated with 0 and 1 degrees of change was significant in the Prediction group ($t = 2.09$; $df = 22$; $p < .05$), and in the No Prediction group ($t = 2.09$; $df = 22$; $p < .05$). The difference between the means associated with 1 and 2 degrees of change was not significant in the Prediction group ($t = 1.44$; $df = 30$; $p > .05$), or in the

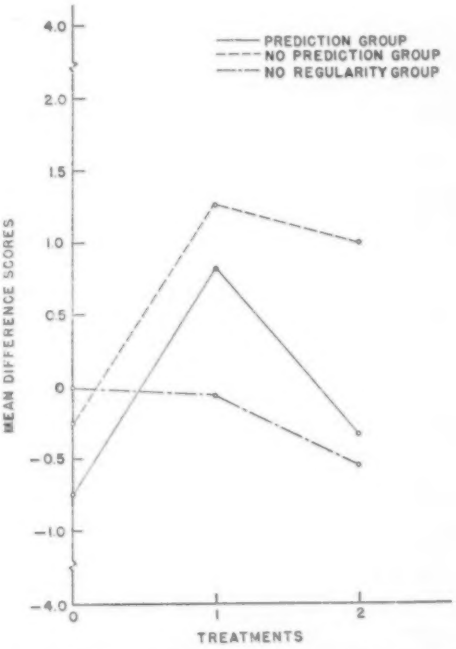

FIG. 2. Mean difference scores associated with Treatments 0, 1, and 2. (The treatments represent, respectively, 0, 1, and 2 degrees of change in order of events for the Prediction and No Prediction groups.)

No Prediction group ($t = 0.33$; $df = 30$; $p > .50$). Although the curves seem somewhat consistent with the second hypothesis, part of the data could have occurred by chance. Possibly, the predicted curve might have been demonstrated more conclusively if greater degrees of change had been included in the experimental design. The interpretation of the results as somewhat consistent with expectation is perhaps strengthened by the lack of curvilinearity shown by the affective change curve of the No Regularity group.

In the Prediction group, 1 degree of change was associated with positive affective change, and with inaccuracy of prediction of 42%. The treatment constituting 2 degrees of change was associated with inaccuracy of 64%, and with negative affective change. The no change treatment was associated with inaccuracy of 8%, and with negative affective change. The group's affective change curve can be understood in terms of the proposed effects of increasing levels of series unexpectedness upon affective tone.

Treatment 3, which constituted a change in physical characteristics of the series, was presented to the subjects of Subgroup D in each group. The mean difference scores for Subgroup D in the Prediction, No Prediction, and No Regularity groups were 1.38, 1.25, and −0.25, respectively. Each of these means was compared to the Subgroup A mean in its group. The resulting differences are significant in the case of the first mean ($t = 2.26$; $df = 14$; $p < .05$), and the second mean ($t = 2.17$; $df = 14$; $p < .05$), but not in the case the last mean ($t = 1.67$; $df = 14$; $p > .10$).

Discussion

The results obtained during environmental regularity are consistent with the proposition that both greatly unexpected and completely expected situations arouse negative affect, while moderately unexpected situations arouse positive affect. The results obtained during environmental change are less clearly in support of this position. Additional investigation involving a wider range of change would be useful.

An alternative explanation of the results of this experiment would involve the assumption that the task of completing sentences may have initially been experienced as threatening, and therefore unpleasant. The individual may have feared that he would not function adequately, or that he would disclose embarrassing aspects of his personality. Some experience with the task may have led to a feeling of mastery, and even enjoyment. After completing even more stems, the task may have become tedious, and no longer challenging. While such a position would account for the affect curves of the Prediction and No Prediction groups during regularity, it does not explain the positive affective change that was associated in these groups with 1 degree of change. Presumably, completing the final stem was no more challenging than completing the stem before the last. In addition, if such a position were tenable, the No Regularity group's affect curve during the first eight series should have been similar to the curves of the other two groups.

It will be recognized that the correct and incorrect predictions made by the Prediction group can be considered mild rewards and punishments, respectively. During environmental regularity, the frequency of rewards increased, while that of punishments decreased. That the curve of affect did not remain in the positive range once having become positive, indicates the difficulty of explaining the results of this experiment in terms of reinforcement theory. Of course, the rewards and punishments were probably very mild. If accurate prediction had been more important to the subjects, the group curve might not have become negative again after so few repetitions of the series.

Although it has been assumed that the effects of the unexpectedness of the situation are most clear when rewards and punishments, as usually conceptualized, are at a minimum, it need not be concluded that the unexpectedness factor is trivial in determining affective tone. Results reported by Spector (1956) suggest that the positive affective tone produced by a rewarding event can be enhanced if the event is also relatively unexpected. Using a procedure in which small groups were given a task to perform, he found that among the subjects who received "promotions," those who had perceived this rewarding event as low in probability responded with higher morale than did those who had perceived it as high in probability. The nature of the interaction

between the affect produced by the unexpectedness of the situation and by its reward value deserves further study.

SUMMARY

This experiment concerned affective tone during environmental regularity and change. Two groups experienced environmental regularity, which consisted of eight presentations of an initially novel series of events. Following this, the various subgroups within each group experienced different versions of a final series of events. Each of the versions constituted a different degree of change in order of events. A control group received the same number of series as did the other two groups, but did not experience environmental regularity in that the order of events in each successive series was different from that of the preceding series.

The last stimulus in each series was a sentence completion stem. As there were nine series in all, each subject perceived nine sentence stems. The subjects were instructed to complete the stems whenever they encountered them in the course of the experiment. The completions were used as indicants of affective tone.

The results indicated that as the series was repeated (environmental regularity), affective tone was initially negative, became positive, and then became negative again. The clearest finding concerning the effects of change in order of events was that the relatively moderate degree of change was associated with affective change in the positive direction. The results are somewhat consistent with the proposition that greatly unexpected and completely expected situations arouse negative affect, while moderately unexpected situations arouse positive affect.

REFERENCES

BINDRA, D. Stimulus change, reactions to novelty, and response decrement. *Psychol. Rev.*, 1959, **66**, 96–103.

BOUSFIELD, W. A. The relationship between mood and the production of affectively toned associates. *J. gen. Psychol.*, 1950, **42**, 67–85.

FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.

GEORGE, C. E., & BONNEY, W. C. Rorschach's affect-color hypothesis and adaptation level theory. *Psychol. Rev.*, 1956, **63**, 294–298.

HEBB, D. O. Drives and the C. N. S. (conceptual nervous sytem). *Psychol. Rev.*, 1955, **62**, 243–254.

KARSTEN, ANITRA. Psychische sättigung. *Psychol. Forsch.*, 1928, **10**, 142–154.

KELLY, G. A. *The psychology of personal constructs*. Vol. 1. *A theory of personality*. New York: Norton, 1955.

KENNY, D. T. The contingency of humor appreciation on the stimulus-confirmation of joke-ending expectations. *J. abnorm. soc. Psychol.*, 1955, **51**, 644–648.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

MADDI, S. R. Surprise and affective arousal during environmental regularity and change. Unpublished doctoral dissertation, Harvard University, 1959.

SHIPOLA, ELSA M. The influence of color on reactions to ink blots. *J. Pers.*, 1950, **18**, 358–382.

SPECTOR, A. J. Expectations, fulfillment, and morale. *J. abnorm. soc. Psychol.*, 1956, **52**, 51–56.

SULLIVAN, H. S. *The psychiatric interview*. New York: Norton, 1954.

(Received February 29, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 346-351

THE EFFECTS OF FRUSTRATION AND ANXIETY ON OVERT AGGRESSION[1]

JACK E. HOKANSON
Florida State University

This study is aimed at investigating some of the direct, physical resultants of experimentally aroused anger. For this purpose three major variables are considered: the degree of frustration, the retaliatory powers of the frustrator, and subject differences in apparent readiness to express aggression. Each of these variables have been dealt with by Dollard, Doob, Miller, Mowrer, and Sears (1939) in their original formulation of the frustration aggression hypotheses, and the empirical findings with respect to these factors have been recently reviewed by Berkowitz (1958).

While most investigations support the hypotheses put forth in *Frustration and Aggression*, a general limitation of these studies involves the measuring devices employed. Typically, "overt aggression" is measured by means of questionnaire responses, ratings, or interviews concerning feelings of unfriendliness toward the frustrator, and a number of indices representing displaced aggression—e.g., doll play and TAT content. Such questionnaires and ratings are vulnerable to the subject's conscious control over his responses in a socially desirable direction. Further, unfavorable ratings of the experimenter, or aggressive TAT protocols, may express merely a negative evaluation of the frustrator, not "aggression" in the strict sense required by the specification (Dollard et al., 1939) that the goal response of the instigation to aggression involves inflicting injury to an organism.

The present research attempts to overcome these difficulties by using the application of direct, physical pain to the frustrator (electric shocks) as measures of aggression. Furthermore, through the use of a specially designed apparatus, the physical characteristics of a subject's aggressive behavior can be observed along three major dimensions—number of aggressive acts, average duration of aggressive act, and average intensity of aggressive act. This type of approach should provide more direct tests of frustration aggression hypotheses as they apply to overt, physical hostility, and permit a more detailed analysis of the physical dimensions of aggressive behavior.

METHOD

The aim of this study was to establish, in a laboratory setting, some of the conditions affecting the expression of overt, physical aggression. Subjects were presented with the following series of events: They were led to believe that this was an experiment on the physiological concomitants of intellectual performance;[2] informed that if they did not fully cooperate, or displeased the experimenter in any other way, the experimenter could punish them by administering an electric shock (retaliatory anxiety manipulation); degraded and insulted by the experimenter for a presumably poor performance on an intellectual task (frustration manipulation); given an opportunity to hurt the experimenter physically by giving him electric shocks.

The manipulations of anxiety and frustration have control conditions in which the subjects were not so treated, and in the overall design, these two variables are orthogonal to each other. A third, subject variable is orthogonal to these. Subjects were divided into two groups depending on whether they scored low or high on a number of previously administered psychometric measures of hostility (High or Low Test Hostility). Thus there are two levels of each of the three variables arranged in a factorial design: High or Low Test Hostility subjects, high or low retaliatory anxiety manipulation, and high or low frustration.

Subjects

Eighty white male undergraduate students at the University of Wisconsin selected from a pool of 375 males taking an introductory or intermediate course in psychology served as subjects. Subjects were chosen on the basis of their scores on three psychometric instruments: Siegal's (1956) Manifest Hostility Scale (MHS), three TAT protocols (8BM, 3BM, 2) scored for hostile content, and first impression ratings of two photographs in terms of the degree of friendliness the subject felt toward the persons portrayed. Subjects who consistently scored above the median on each test distribution (High Test Hostility subjects) and subjects who consistently scored below the median on each distribution (Low Test Hostility subjects) were considered eligible to participate. These subjects were subsequently

[1] This study was conducted as part of the research program sponsored by Grant M-1540, National Institute of Mental Health, United States Public Health Service, Leonard Berkowitz, principal investigator.

[2] This paper is part of a larger research project involving physiological responses to anger and fear provocation, the details of which are reported elsewhere (Hokanson, 1959).

contacted by phone, with approximately 85% of those contacted agreeing to participate in the research. The selection procedure thus provided 40 Low and 40 High Test Hostility subjects. Details of the selection procedure are available elsewhere (Hokanson, 1959).

Procedure

The experimental session, an individual meeting lasting approximately 20 minutes, was divided into a number of phases. The experimenter introduced the experiment as one involving measurement of the physiological concomitants of routine intellectual work. The subject was then placed in the blood pressure and GSR recording apparatus and allowed to rest for 8 minutes. After the adaptation period the subjects who had been randomly assigned to the high anxiety condition were told that on subsequent tasks it was highly important that they cooperate fully and devote maximum energy to their performance, and that if the experimenter felt they were not doing this he could administer an electric shock to them. (Actually this was never done, but merely served as a threat of physical pain to the subject.) Subjects in the low retaliatory anxiety condition were also told that cooperation and maximum energy were important, but that they would probably do better if they just relaxed and did the best they could.

The first "intellectual task" of this session was then introduced. The subject was asked to count backwards from 100 to 1 by threes as quickly as possible. Subjects assigned to the high frustration condition were then exposed to repeated interruptions, insults, and sarcastic remarks by the experimenter. In the low frustration condition the experimenter omitted these remarks and said "Good" when the subject reached zero. The task took approximately 1 minute.

Ratings of subjective feelings of anxiety and anger were obtained at several points in the procedure in order to check the success of the experimental manipulations. Each scale consisted of three items to which the subject was to respond by placing a check mark at an appropriate point along a 160 centimeter line. Only the end points of each line were anchored with a verbal description of an extreme subjective feeling. A typical item on the anxiety questionnaire was "How keyed up do you feel at present?" The hostility items asked such things as "What were your general feelings toward the experimenter during the experiment?" The subject's score consisted of the sum of the distances his check marks were from the positive end of each line for the three items.

In evaluating the success of the retaliatory anxiety manipulation, the postmanipulation anxiety rating was subtracted from the score obtained during the initial adaptation period. Similarly, in evaluating the success of the frustration manipulation, the postfrustration aggression rating[3] was subtracted from the same rating obtained prior to frustration. The analyses of these data, presented in an earlier report (Hokanson, 1959), indicate that both the retaliatory anxiety and the frustration manipulations had their desired effects.

Measurement of overt aggression. The method for measuring overt, physical aggression was one originally suggested by Hokanson and Gordon (1958) in which the subject had the option of administering electric shocks to the experimenter. Following the manipulation of frustration, the experimenter introduced the situation with the following instructions:

> The next part of this session is something quite different. As you are probably well aware, one of the biggest sources of error in the interpretation of psychological tests is the interpreter himself. For instance, if I were to interpret your tests [taken some 3 weeks earlier as part of the screening battery] such factors as being in a bad mood or feeling physically ill would certainly affect my interpretation. This is the problem I am attempting to investigate in this part of the session. I am going to read out loud some items from a personality questionnaire and try to predict how I think you would answer each item. If my prediction is wrong I want you to give me an electric shock through this apparatus. If I am correct in my prediction, simply don't do anything and I will go on to the next item. In this way I will feel physical discomfort when my predictions are wrong and I would like to see how this affects my subsequent predictions.

The questionnaire was one developed to measure attitudes toward dating at the University of Wisconsin. The items are rather straightforward, innocuous statements that appear to keep personal involvement at a minimum. Thus either to agree or to disagree with a particular item should make relatively little difference to the subject. His decision to shock or not shock should therefore be determined mostly by his readiness to express hostility overtly. The number of times the subject disagreed with the experimenter's prediction (i.e., number of times the subject shocked the experimenter) was thus used as one measure of overt hostile expression.

The questionnaire consisted of 15 items on which at least 80% of the standardization groups of University of Wisconsin male students gave the same response. Each "prediction" was made in a direction opposite from that which the normative population indicated to be the prevailing attitude. In this way, each subject was faced with 15 predictions about himself with which he most likely disagreed. Situated in front of the subject was an elaborate "shock apparatus" with electrodes connected to the experimenter's hand.

The subject was thus given 15 opportunities to administer electric shocks to the experimenter by depressing a plunger situated on top of the apparatus. (The experimenter never actually received a shock, but merely behaved as if he had.) The subject's responses with respect to the plunger were recorded along three major behavioral dimensions: the number of shocks administered, the average duration per shock, and the average maximum pressure exerted per shock.

The "shock" apparatus was constructed especially for this research in order to provide these measures. (The plunger that presumably activated the shock was

[3] This rating was postponed until after the subject had an opportunity to aggress physically against the experimenter. While it would have been preferable to have the subject rate his feelings immediately after frustration, focusing the subject's feelings at that time, and having these feelings expressed to the experimenter indirectly via the rating, might have contaminated the subsequent behavioral measures of aggression.

actually connected directly to a spring scale with a maximum reading of 6 pounds. When the subject depressed the plunger a direct reading from the large face of the scale (visible only to the experimenter) was made and recorded. In addition, the base of the plunger consisted of a small metal contact that acted as one pole of a switch in an electric timer circuit. Depression of the plunger with one fourth pound of pressure closed the circuit, which remained closed until the plunger returned to its resting position. A cumulative record of the time that the subject was actually shocking the experimenter was thus produced.

As part of a larger battery of physiological measures, systolic blood pressure was recorded at several points during the procedure: at the end of the initial adaptation period, after the anxiety manipulation, after frustration, during the expression of aggression to the experimenter, and during a final rest period 1 minute after aggression towards the experimenter. By comparing a particular blood pressure level with the immediately preceding level, systolic changes at each of the experimental steps could be calculated. These measures are included in the present report as informative correlates of the behavioral measures of aggression.

RESULTS

Table 1 presents an analysis of variance of the number of shocks administered to the experimenter. The significant main effect for frustration indicates that the subjects who were in the high frustration condition administered a greater number of shocks to the experimenter than did the subjects in the low frustration condition. Subjects under high frustration administered an average of 9.25 shocks while those under low frustration shocked the experimenter an average of 7.87 times. Also of note in this analysis is the absence of significant differences between the subjects in the high and low retaliatory anxiety conditions and between the subjects High and Low in Test Hostility.

Table 2 presents an analysis of variance of the scores for average duration per shock (subjected to square root transformation). None of the terms in this analysis is significant.

The analysis of variance of the scores for average pressure per shock is shown in Table 3. Note the significant Frustration × Anxiety interaction. The cell means and the results of a Duncan range test, presented at the bottom of Table 3, indicate that the mean of the Low Frustration-Low Anxiety condition (2.78) is just short of being significantly lower than the means in the Low Frustration-High Anxiety (3.78) and High Frustration-Low Anxiety (3.80) conditions ($p = .06$). Thus there is suggestive evidence that under conditions of low retaliatory anxiety, high frustration leads to increased pressure per shock, while under low frustrating conditions, being subject to high retaliatory anxiety also leads to increased pressure per shock. However, under both high frustration and high retaliatory anxiety, the pressure per shock is not significantly different from that occurring under Low Frustration-Low Anxiety.

Table 4 presents a summary of the Pearson product-moment correlations obtained between the three behavioral measures of aggression and the systolic blood pressure and subjective feeling changes of the 40 subjects in the high frustration condition. Individual correlations were not routinely computed for each separate experimental condition because of the small number of cases available. However, where a significant r was obtained for all 40 subjects, individual cell scattergrams were plotted and inspected for possible differences among con-

TABLE 1
NUMBER OF SHOCKS ADMINISTERED TO THE EXPERIMENTER

Source	*df*	*SS*	*MS*	*F*
Test Hostility	1	0.3	0.3	
Frustration	1	37.8	37.8	6.30*
Anxiety	1	4.5	4.5	
Test Hostility × Frustration	1	10.5	10.5	
Test Hostility × Anxiety	1	9.1	9.1	
Frustration × Anxiety	1	20.0	20.0	
Test Hostility × Frustration × Anxiety	1	1.2	1.2	
Residual	72	432.3	6.0	
Total	79	515.7		

* $p < .05$.

TABLE 2
AVERAGE DURATION PER SHOCK ADMINISTERED TO THE EXPERIMENTER (in seconds)

Source	*df*	*SS*	*MS*
Test Hostility	1	.59	.59
Frustration	1	.06	.06
Anxiety	1	.01	.01
Test Hostility × Frustration	1	.70	.70
Test Hostility × Anxiety	1	.15	.15
Frustration × Anxiety	1	.67	.67
Test Hostility × Frustration × Anxiety	1	.31	.31
Residual	72	19.45	.27
Total	79	21.94	

Note.—No significant F was obtained.

TABLE 3
AVERAGE PRESSURE PER SHOCK ADMINISTERED TO THE EXPERIMENTER (in pounds)

Source	*df*	*SS*	*MS*	*F*
Test Hostility	1	.61	.61	
Frustration	1	1.51	1.51	
Anxiety	1	1.30	1.30	
Test Hostility × Frustration	1	.13	.13	
Test Hostility × Anxiety	1	.77	.77	
Frustration × Anxiety	1	11.12	11.12	9.11**
Test Hostility × Frustration × Anxiety	1	1.05	1.05	
Residual	72	87.51	1.22	
Total	79	104.00		

Cell Means in the Frustration × Anxiety Interaction

	Low Anxiety	High Anxiety
Low Frustration	2.78	3.78
High Frustration	3.80	3.31

Note.—There are no significant differences among the cell means by Duncan range test.
** $p < .01$.

TABLE 4
PEARSON PRODUCT-MOMENT CORRELATIONS OF AGGRESSION MEASURES WITH SYSTOLIC BLOOD PRESSURE AND SUBJECTIVE FEELING CHANGES
($N = 40$ subjects in high frustration condition)

Measure	Change during Frustration		Change during Shock the Experimenter	Change during Final Rest
	Anger rating	Systolic	Systolic	Systolic
Number of Shocks	−.23	−.02	−.22	−.13
Duration per Shock	−.03	−.15	+.10	−.19
Pressure per Shock	−.32*	−.38*	+.39*	−.42*

* $p < .05$.

ditions. Note that the four significant correlations all occur with the scores for average pressure per shock, previously thought to be a reliable indicator of a presumed instigation to direct, overt aggression.

The r of −.32 between pressure per shock and changes in subjective ratings of anger appears to be somewhat paradoxical. It will be recalled however, that the postfrustration anger rating was actually obtained after the "Shock the experimenter" portion of the procedure. Thus this negative r indicates that the greater the vigor of the subjects' responses while aggressing against the experimenter, the less they subsequently report having felt angry at the experimenter. Inspection of the four individual condition scattergrams revealed no major differences in this trend among the cells.

The r of −.38 between pressure per shock and systolic changes during frustration requires further elaboration. Because of the obvious differences in the scatterplots of each condition, individual correlations based on an N of 10 in each cell were computed. The correlations for Low and High Test Hostility subjects under low retaliatory anxiety conditions were +.15 and +.03, respectively, while under high anxiety conditions they were −.71 and −.77, respectively. These latter two significant correlations seem to indicate that when there is an objective possibility of physical punishment, the greater the increase in systolic blood pressure (very likely a reliable indicator of anxiety) during frustration, the less the vigor of the subjects' subsequent aggressive behavior.

The positive correlation of .39 found between pressure per shock and systolic increases during "Shock the experimenter" indicates that the greater the vigor of aggressive responses, the greater the increase in systolic blood pressure during the expression of overt aggression. By the same token, the −.42 correlation between pressure per shock and elevation of systolic blood pressure at the final rest period suggests that the more vigorous the overt behavior while shocking the experimenter, the greater the reduction in systolic pressure after the aggressive responses are completed.

DISCUSSION

The present study focuses upon three dimensions of direct, physical aggression: the number of aggressive acts, the duration of such acts, and the intensity of physical exertion during these acts. Apparently, these three aspects are not equally sensitive as indicators of the frustration produced instigation to overt hostility, assuming the validity of the classical analysis by Dollard et al. (1939). While number of aggressive acts did reliably increase with frustration, judging from the number of significant relationships with each measure, the vigor with which the hostile act is carried out seems to be at least an equally sensitive index of this instigation. Frustration

aggression theorists maintain that overt hostility is inhibited by threat of punishment for aggression, but that as itself is frustrating, the blocking of the hostile behavior only further increasing the strength of the instigation to aggression. Consistent with this, there was the strongest pressure per shock (i.e., the most vigorous direct hostility) in the High Frustration-Low Anxiety and Low Frustration-High Anxiety conditions (Table 3). The similar results under these two treatments suggest that retaliatory anxiety did indeed serve to heighten the instigation to overt aggression.

However, if both anticipation of punishment and the experimental frustration increase this instigation, the question arises as to why, under a combination of both high frustration and high retaliatory anxiety, there was not an even greater pressure exerted per response, rather than the reduced pressure actually found. An application of Miller's "approach-avoidance" conflict model suggests that the subjects in the High Frustration-High Anxiety condition may have been so strongly angered that they recognized the danger of the experimenter's retaliation, and therefore inhibited their hostile acts. In other words, the strength of the instigation to aggression had to exceed a certain minimum before the inhibition could take place. This minimum level may not have been passed in the High Frustration-Low Anxiety or the Low Frustration-High Anxiety conditions and the subjects more readily expressed their hostility towards the experimenter.

There are several studies in the literature offering evidence that high levels of aggressive drive can be anxiety arousing. Schachter (1957) found a trend in his data indicating that the subjects rated as being moderately angry after frustration manifested a nor-epinephrine-like physiological response, whereas the subjects rated as being either mildly or intensely angry, showed an anxiety type of physiological reaction (epinephrine-like). Along similar lines, Hokanson (1959) demonstrated that the subjects who admit to strong feelings of anger on personality tests manifest more of an anxiety type of physiological response following frustration than do the subjects who get low aggression scores on the tests.

The study provides further evidence in support of the contention that anxiety results in reduced manifestations of overt hostility. There is ample evidence suggesting that increase in systolic blood pressure is a reflection of anxiety (Ax, 1953; Schachter, 1957); thus the significant negative correlation between systolic increases during frustration and subsequent pressure exerted per shock in the high anxiety condition indicates that the greater the physiological manifestations of anxiety during an instigation to aggression, the less the vigor of postfrustration hostile acts. Another line of indirect evidence pertaining to the reduction in aggressive tendencies, presumably because of the arousal of anxiety, comes from the significant negative correlation between pressure exerted per shock and ratings of anger felt during frustration. It will be recalled that this anger rating was obtained *after* that subject had an opportunity to administer shocks to the experimenter. The negative relationship thus indicates that the greater the vigor of the subject's response while he was aggressing against the frustrator, the less his subsequent rating of how angry he was during frustration, as if the previous anger were now being denied. This finding conforms to Berkowitz' (1958) contention that high intensity of aggressive behavior may provoke guilt and/or aggression-anxiety which in turn tends to reduce subsequent manifestations of aggression.

Of added interest are the correlations between pressure per shock and systolic changes occurring during and after the expression of aggression towards the experimenter. The pattern of changes indicates that the vigor of response is positively correlated with increase in systolic blood pressure during the actual expression of aggression, but negatively correlated with elevation of blood pressure when the aggressive behavior is completed. These findings suggest that while overt aggression may momentarily increase certain physiological activities, it is also associated with a later reduction in those activity levels. Worchel (1957) suggested a similar cathartic-like reduction in "tension" following the expression of aggression in his study of frustration produced interferences with intellectual performance. The present findings are only suggestive, however, owing to the possible operation of

sequence effects and homeostatic controls on the course of blood pressure during the latter part of the experiment.

The overall results of this study suggest that both number of aggressive acts and the vigor with which aggressive behavior is carried out may be reliable indicators of hostile motivation. While neither measure provided entirely consistent data, their potential usefulness in studying the relationship between physiological and behavior processes in the area of aggression is indicated.

SUMMARY

This study investigates some of the behavioral concomitants of experimentally aroused anger. An orthogonal design, with 80 undergraduate males, varied test hostility, level of experimental frustration, and level of retaliatory anxiety, the latter variable being manipulated by leading half of the subjects to believe that the experimenter had the potential for inflicting physical pain to them via electric shocks if they did not cooperate fully during the experiment. Following these experimental manipulations the subjects were given a number of socially sanctioned opportunities to aggress against the experimenter by administering electric shocks to him. The subjects' behavior with respect to the plunger activating the shock was analyzed along three dimensions: the number of shocks administered, the average duration per shock, and the mean pressure exerted while activating the shock.

The major findings in this study are:

1. Subjects who were frustrated administer significantly more shocks to the frustrator (the experimenter) than the subjects in the low frustration condition.

2. Subjects exposed to either high frustration-low retaliatory anxiety or high retaliatory anxiety-low frustration manifest a strong trend toward greater vigor of aggressive responses than the subjects under low frustration-low retaliatory anxiety conditions.

3. Among frustrated subjects under high retaliatory anxiety conditions there is a significant negative correlation between the presumed degree of anxiety aroused during frustration (systolic blood pressure increases) and the vigor of subsequent aggressive behavior.

4. Among frustrated subjects there is also a significant negative correlation between the vigor of aggressive responses and subsequent ratings of the degree of anger felt toward the experimenter during frustration.

REFERENCES

AX, A. F. The physiological differentiation between fear and anger in humans. *Psychosom. Med.*, 1953, **15**, 433–442.

BERKOWITZ, L. The expression and reduction of hostility. *Psychol. Bull.*, 1958, **55**, 257–283.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

HOKANSON, J. E. Some physiological and behavioral concomitants of experimentally aroused anger. Unpublished doctoral dissertation, University of Wisconsin, 1959.

HOKANSON, J. E., & GORDON, J. E. The expression and inhibition of hostility in imaginative and overt behavior. *J. abnorm. soc. Psychol.*, 1958, **57**, 327–333.

SCHACHTER, J. Pain, fear and anger in hypertensives and normotensives: A psychophysiological study. *Psychosom. Med.*, 1957, **19**, 17–29.

SIEGAL, S. M. The relationship of hostility to authoritarianism. *J. abnorm. soc. Psychol.*, 1956, **52**, 368–372.

WORCHEL, P. Catharsis and the relief of hostility. *J. abnorm. soc. Psychol.*, 1957, **55**, 238–243.

(Received March 8, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 352-355

THE PERFORMANCE OF PSYCHOPATHIC INDIVIDUALS UNDER CONDITIONS OF POSITIVE AND NEGATIVE PARTIAL REINFORCEMENT[1]

DONALD H. PAINTING[2]

United States Public Health Service Hospital, Lexington, Kentucky

While there is general agreement among clinical observers regarding the distinctive characteristics of psychopathic behavior, disagreement exists concerning the variables which control this behavior. One group (Cleckley, 1955; Henderson, 1947; Karpman, 1948) interprets the poor judgment and need for immediate gratification of the psychopath as manifestations of an insufficient internalization of the mores of society. Thus, the probability of responses in violation of legal and moral codes is higher as these responses are not followed by anxiety and guilt. A second group (Aichhorn, 1935; Lindner, 1944, 1946; Pennington, 1954) takes an opposing view. The seeming inability of the psychopath to profit from experience (differential rewards and punishments) springs from overly severe standards of self-conduct. According to this latter conception, asocial behavior serves to secure punishment and, hence, to mitigate guilt. Some attempt has been made to reconcile this conflict (Karpman, 1948; McCord & McCord, 1956) by the suggestion that two subdivisions exist within psychopathy, the primary or "ideopathic" type and the secondary or "symptomatic" type, where the basic personality is neurotic in nature.

Unfortunately, relatively little experimental evidence is available concerning these issues. Fairweather (1954), using a rote serial learning method, has demonstrated that psychopathic criminals do not perform in an inferior fashion under appropriate motivating conditions. Doubt is therefore cast on the assertion that psychopathy is invariably accompanied by a basic deficiency in learning capacity. Experimental support for the poor judgment observed clinically in the psychopath has been obtained, however. A motor skill task was employed by Lonstein (1952) in a study using level of aspiration as the dependent variable. The goal setting behavior of the psychopath was found to be unrealistic and largely unmodified by success or failure in the task.

The present investigation was a further effort to explore the validity of clinical observation in a more controlled environment. A two-choice partial reinforcement procedure was used, both because of the greater similarity to naturally occurring reinforcement schedules (Ferster & Skinner, 1957) and the comment by many writers concerning the abnormal risk taking behavior of psychopathic individuals (Noyes, 1955). Negative as well as positive reinforcement was used to determine whether the behavior of psychopathic individuals was differentially affected by a correct response's being followed by the attainment of a reinforcing stimulus or the avoidance of an aversive one. Three probability learning sequences, in which a correct response on trial (n) was determined by a preceding stimulus of successively greater distance from trial (n), were employed. This served to provide a test of the contention that the psychopath is less capable of perceiving the relationships between past events and the consequences of present actions.

Method

Subjects

On the basis of internalization ratios (IR) (Welsh, 1952) derived from the MMPI (Hathaway & McKinley, 1951), two experimental groups of 36 subjects each were selected from among Negro male postnarcotic drug addicts hospitalized at the United States Public Health Service Hospital, Lexington, Kentucky. The IR was designed to indicate the direction in which anxiety is expressed, either toward the self, IR $>$ 1.00, or toward the environment, IR $<$ 1.00. One group, the primary psychopaths (PP), obtained a mean IR of 0.79, while the other group, the neurotic psychopaths (NP), received a mean IR of 1.02. By this device, two psychopathic groups (both had MMPI *Pd* scores greater than 70) were formed which appeared to cor-

[1] This study is part of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy to the Graduate School of the University of Kentucky.

Appreciation is gratefully expressed to John W. Donahoe for his critical suggestions and helpful assistance in this project.

[2] Now at Veterans Administration Hospital, Coatesville, Pennsylvania.

TABLE 1
MEAN *T* SCORES ON MMPI SCALES

	L	*F*	*K*	*Hs*	*D*	*Hy*	*Pd*	*Mf*	*Pa*	*Pt*	*Sc*	*Ma*
Normal	46	53	53	52	53	53	60	53	47	56	59	65
NP	50	58	53	65	77	64	79	63	59	69	63	63
PP	53	53	57	49	56	56	74	59	50	50	53	65

respond to the suggestion made by Karpman (1948) and McCord and McCord (1956) regarding the subdivisions within psychopathy. The complete MMPI scores for all groups are presented in Table 1. The control (C) group, consisting of 36 Negro male college students, had a mean IR of 0.90. The mean of C differed by one standard deviation from the means of both psychopathic groups, indicating a minimum of overlap. Shipley-Hartford vocabulary scores (Shipley, 1940) were obtained from all subjects to insure that differences in treatment effects did not arise from original differences in intellectual ability. The mean vocabulary scores were 26.08, 26.53, and 26.75 for PP, NP, and C, respectively. In no comparison was there a statistically significant difference.

Apparatus

The apparatus used in the experiment was similar to that employed by Grant, Hake, and Hornseth (1951). Two white lights were located to the left and right of the center of a horizontal panel. The subject sat facing the lights, while the experimenter sat behind the panel, inaccessible to the subject's vision. A small red signal light located between the white lights served to alert the subject before each trial. Two push-button switches were mounted on the base of the panel, one below each of the white lights. A hinged door and a shelf were placed in the center of the panel below the signal light.

Procedure

In order to determine which response alternative, right (R) or left (L), was correct on a given trial, three sequences were constructed according to a method described by Smith (1955). In Sequence I, the correct alternative on trial (n) was independent of all preceding trials. In Sequence II, an alternative repeated itself with $p = .25$ and changed to the other alternative with $p = .75$. Thus, trial (n) was dependent on trial (n-1). For Sequence III, trial (n) was independent of trial (n-1) but dependent on trial (n-2). That is, if R was correct on trial (n-2), R was correct on trial (n) with $p = .25$. In all three sequences, the probability of occurrence of both alternatives was .50 in a series of 200 trials.

The three groups (PP, NP, and C) were partitioned into three equal subgroups, each of which received one of the sequence treatments. For one-half of each subgroup, 5 correct responses resulted in a reward of 2 cigarettes. For the other half, 5 incorrect responses resulted in the loss of 2 cigarettes from an original 80 given to the subject at the beginning of the experiment by the experimenter. Poker chips, which could be exchanged for cigarettes at the end of the experiment, were used as the medium of exchange. The chips were given to or taken from the subject through the hinged door in the apparatus. All subjects, after having been told the manner in which cigarettes could be won or lost, were given 200 trials on the appropriate sequence. The response on each trial was recorded.

RESULTS

Sequence I

Performance under Sequence I, in which the conditional probabilities approximated the overall probability of an alternative, served both as a baseline against which Sequences II and III might be compared and as an index of behavior under conditions which precluded a strategy insuring greater than 50% reinforcement. The number of instances in which the subject selected a response opposite to the response which was correct on the preceding trial served as the dependent variable. The median test was used throughout to evaluate the data. No differences were found between groups under positive reinforcement; however, in the condition in which a correct response served to avoid the loss of a cigarette, significant differences between groups emerged ($p < .02$). Whereas PP made fewer alternations than either NP or C ($p < .01$ in both cases), NP and C did not differ significantly from one another (see Table 2).

Sequence II

Optimum performance on Sequence II required that the subject learn to choose that response which was incorrect on the preceding trial. Under avoidance conditions, all groups approximated the proportion of actual alternation (.75) in the last 100 trials (see Table 2). Under the reward condition, group differences were observed ($p < .05$). Again, NP and C did not differ significantly, while PP alternated *more* frequently than did NP ($p < .01$) and C ($p < .05$). Thus, PP alternated in excess of the actual conditional probability, a strategy superior to that adopted by either of the other groups.

TABLE 2
PROPORTION OF ALTERNATED RESPONSES DURING THE FINAL 100 TRIALS OF THE SEQUENCE

	Reward			Avoidance		
	PP	NP	C	PP	NP	C
Sequence I	.50	.38	.42	.33	.49	.47
Sequence II	.80	.69	.74	.75	.72	.75
Sequence III	.67	.73	.76	.64	.67	.73

TABLE 3
CHI SQUARE VALUES OBTAINED FROM THE MEDIAN TEST COMPARING PERFORMANCE ON SEQUENCE II AND SEQUENCE III

	Reward		Avoidance	
	chi square	p	chi square	p
NP	2.23	$<.20$	2.27	$<.20$
PP	8.82	$<.01$	4.74	$<.05$
Control	3.44	$<.10$	0.99	$<.50$

Sequence III

In Sequence III, adequate performance required the subject to alternate his responses on trial (n) in accordance with the correct response on trial (n-2). Group differences verged on significance under avoidance ($p < .10$) and attained significance under reward ($p < .02$). The difference between the extreme groups, PP and C, was significant ($p < .01$). Here, PP alternated *less* than did C. When performance on Sequences II and III were compared, only PP showed a significant decrease in the frequency of alternation, the performance of NP and C remaining on the same level in the more difficult sequence (see Table 3). PP alternated less often in Sequence III under both reward ($p < .01$) and avoidance ($p < .05$). In short, PP performance deteriorated generally under conditions necessitating a temporally remote determination of response.

DISCUSSION

The data clearly demonstrate that psychopathy, when defined solely in terms of a deviant *Pd* score on the MMPI, is not a homogeneous behavioral entity. Rather, a subdivision of this group on the basis of the Welsh IR led to the formation of two subgroups, neurotic psychopaths (IR > 1.00) and primary psychopaths (IR < 1.00), whose responses in the partial reinforcement situation were quite different. With manipulation of sequence and reinforcement conditions, the performance of the neurotic psychopath was indistinguishable from that of the normal, whereas the behavior of the primary psychopath was significantly affected by both treatment variables. This finding is compatible with the Karpman (1948) and McCord and McCord (1956) hypothesis and is supported by the work of Hill, Haertzen, and Glaser (1960) and Painting (1954) with the MMPI.

The frequency with which the primary psychopath alternated his responses under the various experimental conditions provides an interesting parallel to clinical observation. Where the reinforcement schedule disallowed any regularity in stimulus-response contingency (Sequence I), the primary psychopath group tended to repeat that response which had been correct on the preceding trial. To the extent that the alternation measure reflects an attempt to secure an appropriate strategy, the behavior of the primary psychopath became stereotyped and rigid in displaying a strong positive recency effect.

With a schedule permitting the control of reinforcement by responses (Sequence II), all groups approximated the ($n = 1$) dependency with the primary psychopath tending to maximize his gains in the reward condition. It is relevant to note that while normals generally match event probabilities, maximizing behavior has been demonstrated with an experimentally induced "gambling set" (Goodnow, 1955). These data suggest that the gambling set is the customary attitude of the primary psychopath, an interpretation which is in agreement with that of many clinical observers. Notice that the performance in Sequence II is not at all compatible with the position that the psychopath is incapable of profiting from experience. If anything, when the stimulus and response are in close temporal proximity, the level of psychopathic performance is superior to that of normals.

Under a more remote stimulus-response correlation (Sequence III), the performance of the primary psychopath deteriorates, however. This deterioration tends to be general and not restricted to a particular reinforcement operation. It would appear, therefore, that with remote determination of the correct response,

the most common naturally occurring relationship, the primary psychopath is, in truth, less proficient than the normal. The origins and stability of this deficiency under different experimental conditions are, of course, unknown. These findings do indicate, however, the danger in generalizations concerning the psychopath which do not take into account the heterogeneity of the group as conventionally defined and situational variables such as response consequences and contingencies.

SUMMARY

A total of 108 subjects (36 primary psychopaths, 36 neurotic psychopaths, and 36 normals) were studied with a two-choice partial reinforcement method. Three probability learning sequences of differing conditional probabilities were used: independent, ($n = 1$) dependent, and ($n = 2$) dependent. These sequences were used in combination with two response consequences, reward and avoidance. In all comparisons, the behavior of the neurotic psychopaths was indistinguishable from that of the normals. Performance of the primary psychopaths, however, tended to become inflexible in the independent sequence under avoidance conditions and to deteriorate generally with the more remote stimulus-response contingencies. The implications of these findings for statements based upon clinical observation were discussed.

REFERENCES

AICHHORN, A. *Wayward youth.* New York: Viking, 1935.

CLECKLEY, H. *The mask of sanity.* (3rd ed.) St. Louis: C. V. Mosby, 1955.

FAIRWEATHER, G. W. The effect of selected incentive conditions on the performance of psychopathic, neurotic, and normal criminals in a serial rote learning situation. *Dissertation Abstr.*, 1954, **14,** 394–395.

FERSTER, C. B., & SKINNER, B. F. *Schedules of reinforcement.* New York: Appleton-Century, 1957.

GOODNOW, JACQUELINE J. Determinants of choice-distribution in two-choice situations. *Amer. J. Psychol.*, 1955, **68,** 106–116.

GRANT, D. A., HAKE, H. W., & HORNSETH, J. P. Acquisition and extinction of a verbal conditioned response with differing percentages of reinforcement. *J. exp. Psychol.*, 1951, **42,** 1–5.

HATHAWAY, S. R., & MCKINLEY, J. C. *Minnesota Multiphasic Personality Inventory; manual.* (Rev.) New York: Psychological Corporation, 1951.

HENDERSON, D. K. *Psychopathic states.* New York: Norton, 1947.

HILL, H. E., HAERTZEN, C. A., & GLASER, R. Personality characteristics of narcotic addicts as indicated by the MMPI. *J. gen. Psychol.*, 1960, **62,** 127–139.

KARPMAN, B. Conscience in the psychopath: Another version. *Amer. J. Orthopsychiat.*, 1948, **18,** 455–491.

LINDNER, R. M. *Rebel without a cause.* New York: Grune & Stratton, 1944.

LINDNER, R. M. *Stone walls and men.* New York: Odyssey Press, 1946.

LONSTEIN, M. A. A comparative study of level of aspiration variables in neurotic, psychopathic, and normal subjects. Unpublished doctoral dissertation, University of Kentucky, 1952.

MCCORD, W., & MCCORD, J. *Psychopathy and delinquency.* New York: Grune & Stratton, 1956.

NOYES, A. P. *Modern clinical psychiatry.* (4th ed.) Philadelphia: W. B. Saunders, 1955.

PAINTING, D. H. An application of the MMPI to a population of narcotic drug addicts: An exploratory attempt to predict the duration of volunteer hospital confinement. Unpublished master's thesis, University of Kentucky, 1954.

PENNINGTON, L. A. Criminal and psychopathic behavior. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology.* (2nd ed.) New York: Ronald, 1954.

SHIPLEY, W. C. A self-administering scale for measuring intellectual impairment and deterioration. *J. Psychol.*, 1940, **9,** 371–377.

SMITH, J. E. K. Statistical structure in probability learning. *Dissertation Abstr.*, 1955, **15,** 634–635.

WELSH, G. S. An anxiety index and an internalization ratio for the MMPI. *J. consult. Psychol.*, 1952, **16,** 65–72.

(Received March 16, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 356-363

ANXIETY, FEAR, AND SOCIAL AFFILIATION

IRVING SARNOFF AND PHILIP G. ZIMBARDO[1]

Western Reserve University *New York University*

In his recent monograph, Schachter (1959) reports that anticipated exposure to a painful external stimulus determines the degree to which persons wish to affiliate with each other: the greater the anticipated pain, the stronger the desire to await the onset of that pain in the company of others in the same predicament. In attempting to account theoretically for this finding, Schachter mentions such motivational forces as the subjects' needs for reassurance, distraction, escape, and information. However, among the various possible explanations, Schachter appears to favor one derived from Festinger's (1954) theory of social comparison processes. Adapting that theory to the phenomena under investigation, Schachter postulates that the arousal of any strong emotion evokes a need for comparison. Emotions are assumed to be quite unspecific states of affect. Hence, persons can only evaluate the quality, intensity, and appropriateness of their emotions properly by comparing their own reactions with those of others. Moreover, novel emotion producing stimuli should induce a greater tendency to affiliate than familiar stimuli. By definition, a novel stimulus is one that is more difficult to fit into a person's established frame of reference for emotive states. Accordingly, the individual is more obliged to seek out others in order to define the emotional effects of novel stimuli.

The explication of Schachter's (1959) results in terms of the theory of social comparison processes is appealingly parsimonious. However, it requires the assumption that *all* emotive states have the same effect on affiliative behavior. Thus, Schachter, like many contemporary psychologists, does not deal with the possible conceptual distinctions between fear and anxiety. Yet, it seems to us that, by adopting an alternative assumption about the psychological properties of emotions, to be presented briefly below, it is possible to formulate predictions concerning affiliative responses that could not have been derived from the theory of social comparison processes. Indeed, by employing Freud's (1949a, 1949b) conceptual distinctions between fear and anxiety, we are led to predict a tendency toward social isolation—rather than affiliation—as a consequence of certain conditions of emotional arousal.

The present experiment was, thus, undertaken with two objectives: to assess the empirical validity of conceptual differentiation between fear and anxiety, and to evaluate the extent to which the theory of social comparison processes may be applied to the relationship between all emotions and affiliative behavior. In order to implement these objectives, we have conducted an experimental investigation of the differential effects of fear and anxiety upon social affiliation.

Functional Relationship between Emotions and Motives

The guiding assumption of our experiment holds that all emotions are consciously experienced epiphenomena of motives.[2] When our motives are aroused, we experience subjective reactions to which we learn, over time, to attach commonly agreed upon labels that signify the various emotions.

Motive, on the other hand, is defined as a tension producing stimulus that provokes behavior designed to reduce the tension. Each of our motives (innate or learned) requires the

[1] The authors are indebted to the Yale Communication Research Project, directed by C. I. Hovland, for its material support of this study, and to C. I. Hovland for his encouragement and helpful suggestions. Our thanks are extended to Jacob Rabbie and Harold Gerard for their contributions to the study. We also wish to acknowledge the expert and enthusiastic participation of our research assistants, Ted Sheldon, Sally Whitcher, and Ira Grushow.

[2] The concept of motivation which we have chosen to employ has been elaborated elsewhere (Sarnoff, 1960a).

performance of a *different* response for the maximal reduction of its tension.

Fear and Anxiety Viewed as Motives

The motive of fear (which Freud called objective anxiety) is aroused whenever persons are confronted by an external object or event that is inherently dangerous and likely to produce pain. Only one type of overt[3] response can maximally reduce our fear: separation from the threatening aspects of the feared object, accomplished by flight from the object, at one extreme, and conquest, at the other. In the case of fear, then, one's energies are mobilized toward dealing with the external stimulus; to eliminate, through some mode of escape or attack, the threat that is clearly and objectively present in the stimulus.

If we examine the consequences of anxiety (which Freud termed neurotic anxiety), we see no such correspondence between the internal disturbance of the person and an objectively harmful environmental stimulus. Instead, anxiety is typically aroused by stimuli which, objectively considered, are *innocuous*.[4] For example, in the case of the classical phobias, harmless objects possess a special motivational significance for certain people. These objects activate some motive other than fear, and this other motive, in turn, arouses the consciously perceived motive of anxiety. Hence, the emotional reaction of the anxious person is inappropriate to the inherent characteristics of the external stimulus.

Regardless of their content, the motives whose arousal evokes anxiety share a common property: they are all *repressed*. These repressed motives continue unconsciously to press for the reduction of their tensions; and anxiety signals the threat of possible expression of these repressed motives. Consequently, the person develops a number of additional ego defenses that function to safeguard the initial effects of repression. If the ego defenses do their work effectively, the motives are kept under repression, the inner danger passes and the individual's anxiety is reduced.

Implications of the Motives of Anxiety and Fear for Affiliative Behavior

It follows from the foregoing discussion that, when their anxieties are aroused, people are more inclined to become preoccupied with the reassertion of inner self-control than with modes of dealing with the anxiety evoking external object. Because the anxious person tends to be aware of the element of *inappropriateness* in his feelings, he is loath to communicate his anxieties to others. To avoid being ridiculed or censured, he conceals anxiety aroused by stimuli which he guesses do not have a similar effect upon others, and which, he feels, ought not so to upset him. Thus, when anxiety is aroused, a person should tend to seek isolation from others. On the other hand, when fear is aroused and he is unable to flee from the threatening object, he welcomes the opportunity to affiliate. Since the usual responses to fear, flight and fight, are restricted in the experimental situation, the subject seeks other fear reducing responses. Therefore, the probability of affiliation increases because it mediates fear reduction through the potentiality for catharsis and distraction as well as the emotional comparison offered by interpersonal contact.

We are led, therefore, to the hypothesis that the motives of fear and anxiety should influence social affiliation behavior differently: the greater the fear aroused, the more the subjects should choose to be together with others while they await actual contact with the fear arousing object. Conversely, the greater the anxiety elicited, the more the subjects should choose to be alone while they await contact with the anxiety arousing object.

METHOD

The experiment was presented to the subjects as a physiological investigation of the cutaneous sensitivity

[3] Space limitations do not permit a consideration of the two types of covert (ego defensive) responses, denial and identification with the aggressor, which persons may employ in their efforts to cope with external threat. A full discussion of these ego defenses is presented by Sarnoff (1960a).

[4] In fact, since anxiety arousing stimuli are often related to unconscious libidinal motives, they may be regarded by most people as intrinsically pleasurable, rather than in any way painful. For example, owing to the manner in which their heterosexual motives have been socialized, some men may tend severely to repress their sexual cravings for women. Hence, when such men are shown photographs of voluptuous nudes, stimuli which might be quite evocative of pleasurable fantasies among most of their fellows, they are likely to experience anxiety (Sarnoff & Corwin, 1959).

of various parts of the body. A 2 × 2 design was used in which two levels of fear and of anxiety were experimentally aroused. The dependent variable of social affiliation was measured by having the subjects state whether they preferred to spend an anticipated waiting period alone or in the company of others.

Subjects

The subjects were 72 unpaid, male undergraduate volunteers from six introductory psychology classes in Yale University. An additional 36 subjects were used to pretest the manipulations and measuring devices, and an additional 13 subjects were excluded from the analyses because they did not qualify as acceptable subjects, i.e., were friends, misunderstood the instructions, did not believe the rationale.

Procedure

Background information was collected by an accomplice alleged to be from the counseling program of the Student Health Department. A questionnaire was designed to obtain background information on the subjects and also their preferred mode of defense mechanism. The latter data were in response to four Blacky cards. As in a recent experiment by Sarnoff (1960b), each card was accompanied by three alternatives that were to be rank ordered according to the subjects' reaction to the theme of the card (sibling rivalry, achievement, and two of sucking). The alternatives reflected predominantly an acceptance of the motive, projection of the motive upon others, or a reaction formation against the motive.

About one month later, the experimenter was introduced to the psychology classes as a physiological psychologist studying physiological responses to sensory stimuli. The subjects were subsequently recruited individually, and randomly assigned to the four experimental treatments. The specious purpose of the experiment and of the conditions of waiting were further established by marking the experimental room "Sensory Physiology Laboratory"and two nearby rooms "Waiting Room A" and "Waiting Room T." Because of absentees, the size of the groups tested varied from three to five, and was usually composed of four subjects. In order to avoid the development of superficial friendships during the experiment, and eliminate the possibility that the subjects might react to cues from each other or from the experimenter, the subjects were isolated in adjacent cubicles, no communication was allowed, and the tape-recorded instructions were presented through earphones.

The experimental conditions and instructions common to all subjects will be presented first. After rolling up their sleeves, removing their watches from their wrists, and gum or cigarettes from their mouths ("They interfere with the recording electrodes"), the subjects were told:

> Our experiment falls in the general area of physiological psychology. As you may know, one branch of physiological psychology is concerned with the reactions of the sense organs to various kinds of stimulation. Our present experiment deals with the skin [or mouth] as an organ of sensation. We are interested in studying individual differences in response to particular stimuli applied to it.
>
> There has been a good deal of controversy about the relative sensitivity of the fingertips [lips] as compared to the palms [tongue], and upper surface of the hand [palate]. Our experiment will help to provide data upon which we may be able ultimately to draw a detailed map of the cutaneous sensitivity of the human hand [mouth].
>
> In order to measure your physiological reactions, we are now going to attach some instruments to your arm and finger [corner of your mouth]. These instruments are electrodes which are connected to a machine which records exactly the strength of your response to each stimulus. . . . Electrode jelly will be applied first to the area to insure that we get a good electrical contact. (The electrodes were then attached by a female laboratory assistant of middle age.)

In order to provide a reasonable basis for asking the subjects to wait in other rooms (and, thus, for making the choice of affiliation or isolation), the subjects were told that it was necessary to assess their basal rates of responding prior to the application of the actual stimuli. They were led to believe that their individual sensitivities were being recorded while they viewed a series of slides of a typical subject who had participated in the experiment. They anticipated that a waiting period would come after the slides, and then in the second—and purportedly major—part of the experiment their direct reactions to the actual stimuli would be measured. Accordingly, they were told:

> Now that your basal rates have been recorded on our polygraph recorder, it will take us about 10 minutes while we tally the data and reset our measuring instruments so that they will be geared to your individual basal rates as you are run one at a time through the rest of the experiment. While we are doing these things, we are going to ask you to wait in other rooms which are available to us. We will come to get you when it is your turn to go through with the experiment. Incidentally, we have found that some of our subjects prefer to do their waiting alone, while others prefer to wait together with other subjects. Therefore, we are going to give you your choice of waiting alone or with others. In either case, you will be ushered to a comfortable room furnished with adequate reading material.

After indicating their preference of waiting alone or together with others, the subjects also indicated the intensity of this preference on an "open-ended" scale in which 0 represented a very weak preference and 100 a very strong preference. On this relatively unstructured scale there was as much as 175 points of difference between subjects (from "75-alone" to "100-together").

Presentation of the slides during the experiment served two purposes in addition to the one previously mentioned. The content of the slides (appropriate to each experimental treatment) served to reinforce the subjects' differential expectations of the nature and severity of the stimulus situation. Furthermore, the subject seen in the slides became a focal point for measuring the effectiveness of the experimental manipulations. It was assumed that a direct attempt (by means of a scaled question) to appraise the level of the subjects' fear or anxiety would be likely to: sensitize them to the true purpose of the experiment; yield

unreliable results since the subjects might neither be consciously aware of, nor able to verbalize, their anxiety reaction; and evoke resistance since some subjects might not want to admit to being anxious or fearful, calling their masculinity into question.

Therefore, it was necessary to use an indirect, disguised measure to evaluate whether the experimental inductions had actually aroused two levels of both fear and anxiety. Immediately after the slides had been shown (but before the affiliation choices had been made), the subjects were told:

> As you may know, an individual shows his physiological reaction in a variety of behavioral forms. We are interested in seeing whether it is possible to estimate how ill-at-ease or upset individuals are at the prospect of receiving the stimulation in this experiment. Recalling the subject whom you just saw in the slides, how upset or ill-at-ease did he seem to you? Please assign a number anywhere from zero to 100 to indicate your feeling. (Zero = unconcerned, at ease; 100 = extremely concerned and ill-at-ease.)

Since the subject in the slides was a posed model instructed to remain poker faced throughout, it was assumed that there was no objective difference in his expression. Thus, any systematic difference in ratings between groups should reflect a projection of the subjects' own motives upon this screen.

However, because the content of the slides was not identical for every group but rather "tailored" to each specific treatment, it was possible that the model may have actually looked more fearful in the slides shown to the subjects in the High Fear than in the Low Fear condition. As a control check on this possibility, four additional introductory psychology classes (N = 108) served as judges. They were told that the slides were of a typical subject in a recently completed experiment, and their task was to estimate how ill-at-ease and concerned he appeared (on the same scale used by the experimental subjects). Two of the classes saw only the face of the model (the rest of the slide was blacked out) and were told only that he was a subject in a physiological experiment in which stimuli were applied and responses measured. The other two classes saw the entire stimulus field of the slides and were given the same complete description that the experimental subjects received. Since each class of judges rated the slides for all four experimental treatments, the order of presentation was counterbalanced.

After the projective measure of motive arousal and the measure of affiliation, the electrodes were removed and a measure taken of the subjects' reasons for choosing to affiliate or be isolated. This was done with the rationale that a social psychologist had become interested in the fact that some of our subjects preferred to be together while others preferred to be alone, and he had asked us to get some information for him about the reasons underlying this preference.

The questionnaire, designed by Gerard and Rabbie (1960), contained both open-ended and structured questions asking for reasons for the affiliation choice. Finally, the subjects noted whether or not they wished to continue in the experiment. Only one subject (in the High Fear condition) refused to remain for the "stimulation" part of the experiment.

The true purpose, hypothesis, design, and reasons for the various deceptions (and, at a later time, the results) were explained fully to each subject.

High Fear

A high level of fear was induced by leading the subjects to anticipate a series of painful electrical shocks. Although they expected to endure each of the shocks for 2 minutes, the subjects were assured that the shocks would not cause damage or injury.

The female assistant (dressed in a white lab coat, as was the experimenter) then attached electrodes to each subject's arm and fingertip and strapped his arm onto a cotton-padded board. The leads from the electrodes appeared to go to a polygraph recorder, which also was seen in the series of slides of the typical subjects. Another slide showed an enormous electrical stimulator, and the implication was that it was behind a curtain in the experimental room. It was called to the subjects' attention that:

> The four dials shown in the upper right-hand corner of the stimulator enable us to regulate automatically the frequency, duration, delay, and intensity of the shock you will get.

The other slides portrayed the subject with earphones and electrodes attached (like the subjects themselves), "listening to the instructions", and then "about to receive his first painful shock," administered by the experimenter, who could be seen in the background manipulating the dials on the stimulator. A final situational factor that may have enhanced the effectiveness of the High Fear manipulation was that the experimental room housed electrical generators which made a continuous buzzing sound, a cue interpreted by the High Fear subjects as the electrical stimulator "warming up," but unnoticed or quickly adapted to by the other subjects. An unobtrusively posted sign reading "Danger/High Voltage," present only for the High Fear subjects, gave further credence to this notion.

Low Fear

In the Low Fear condition the word "shock" was never used, and all cues in the situation associated with shock, fear, or pain were removed; i.e., no white lab coats, arms not strapped to boards, etc. The expectations of these subjects were guided by instructions stating that our methodology was to apply a 10-second stimulus of very low intensity that would be just sufficient to elicit a measurable physiological response.

In the series of slides viewed by these subjects, the imposing electrical stimulator was replaced by a small innocuous looking apparatus (actually a voltmeter), and the experimenter was seen not in the active role as an agent of pain, but in the passive role of recording data from the polygraph recorder.

High Anxiety

Anxiety was manipulated by arousing a motive that was assumed to have been repressed by most of the subjects. In Freudian terminology, the motive might be called "oral libido," a desire to obtain pleasurable gratification by sucking on objects that are clearly related to infantile nursing experiences. The female breast is, of course, the prototype of such objects, but others include nipples, baby bottles, and pacifiers.

Thus, to arouse this oral motive and, hence, the anxiety that should follow its arousal, subjects in the High Anxiety condition were led to believe that they would have to suck on a number of objects commonly associated with infantile oral behavior. They were told that their task would be to suck on these objects for 2 minutes while we recorded their physiological responses from the skin surfaces stimulated by the objects. In clear view in front of the subjects were the following items: numerous baby bottles, oversized nipples, pacifiers, breast shields (nipples women often wear over their breasts while nursing), and lollipops.

The same variety of stimulus objects was shown arrayed in front of the subject in the slides. He could be seen, tongue hanging out, lips puckered, about to suck his thumb (as one of the objects of stimulation) or one of the other objects. Subjects were told that the contact taped to the mouth recorded the direct reaction to the oral stimulation, while the arm contact recorded peripheral reactions.

Low Anxiety

The instructions to the Low Anxiety subjects did not mention "suck," nor any stimulation that they would receive from putting the objects in their mouths. Moreover, they were led to believe that they would keep each object in their mouths for only 10 seconds. The stimulus objects were not in immediate proximity to the subjects while their electrodes were being attached. The stimulus objects which they anticipated putting in their mouths were shown in the slides: whistles, balloons, "kazoos," and pipes. Since these objects do not require sucking (but rather, in general, blowing), the model's tongue was not seen as he prepared to use the stimuli in the slides.

RESULTS

Evidence of the Effectiveness of the Experimental Manipulations

In using the subjects' estimates of the degree to which the model seen in the slides was upset by the prospect of receiving the stimulation in the experiment, it was assumed that the subjects would tend to project their induced level of fear and anxiety. Table 1, which presents the mean projection scores for each experimental treatment, offers evidence that this assumption was valid and the manipulations effective. The High Arousal subjects perceived the model to be significantly[5] more upset, concerned, and ill-at-ease than did the Low Arousal subjects.

Our theoretical distinction between fear and anxiety, and the way these concepts were operationally defined in this experiment, lead to the prediction that, assuming similarity of past experience, persons facing the same clearly, objectively present threat should react in a relatively homogeneous fashion. This close correspondence between stimulus and response is not assumed to hold for anxiety. We have already noted that a stimulus that produces anxiety for some persons is not an anxiety producing cue for many others. Since the significance of the stimulus depends upon its symbolic and generally idiosyncratic associations, one would expect that a stimulus which elicited anxiety for persons with relevant predispositions (repressed motives) would have less effect on those who had more adequately resolved the conflict over the expression of the same motives. Thus, one way of determining whether our experimental manipulations produced two different motives, fear and anxiety (rather than only two levels of one motive), is to compare the variability in response between treatments.

TABLE 1
MEAN PROJECTION SCORES FOR EACH EXPERIMENTAL TREATMENT

Motive	Level of Arousal		p value
	Low	High	
Fear	24	42	$< .01$ ($t = 3.05$)
Anxiety	14	31[a]	$< .01$ ($t = 2.95$)
	NS	*NS*	

Note.—The larger the score, the greater the degree of projection.

[a] Variance greater than in High Fear group, $p < .10$; *SD* for High Anxiety = 24, for High Fear = 16.

The heterogeniety of response in the High Anxiety group is, as predicted, greater than in the High Fear and the Low Arousal conditions. The same difference in response variability between the High Anxiety group and all other groups is manifested as well in the dependent variable of social affiliation. The questionnaire data to be presented in a later section offer further support to the distinction between fear and anxiety.

Before presenting the major results, it is necessary to account for two possible sources of artifact in the just reported data on projection. They are: by chance sampling, the High Arousal groups could have contained more subjects who characteristically used projection as a mechanism of defense than the Low Arousal groups; and the subject seen in the High Fear and High Anxiety slides was ob-

[5] All p values reported throughout the paper are based on two-tailed tests of significance.

jectively more upset and concerned than he was in the Low Fear and Low Anxiety slides. If either of these alternatives were true, then the projection measure would not be a reflection of differences due to the experimental arousal of levels of fear and anxiety.

The pretest data of the subjects' mode of defense preference on the Blacky Projection test show no initial significant difference between any of the groups in their tendency to use projection.

Among the groups of neutral judges who evaluated all the slides shown in the study, from 68%–98% reported perceiving either no difference in the degree to which the model appeared upset, or a difference opposite to that reported in Table 1. This result holds for both fear and anxiety, and regardless of the order of presentation or amount of the stimulus field seen (model's face only or entire slide). Thus, it appears that the projection measure can be used as an index of the efficacy of the experimental conditions and manipulations.

Effects of Fear and Anxiety on Social Affiliation

The results bearing upon the hypothesis of the study are presented in Table 2, where for each condition the mean intensity of desire to affiliate, as well as the number of subjects choosing to affiliate and to be alone, are presented. It is evident that there is a strong positive relationship between fear and the index of affiliative tendency, but a strong negative relationship between anxiety and affiliation, so that as fear increases affiliation also increases, while an increase in anxiety results in a decrease in affiliation. Thus, our prediction of an interaction between kind of motive and level of arousal is clearly supported by the data. While some 95% of the High Fear subjects chose the "together" alternative (with more than 0 intensity), only 46% of the High Anxiety subjects chose to wait together. The marked mean difference between these groups in intensity of choice (51.0–8.0) is significant well beyond the .01 level ($t = 3.63$). The large mean difference in affiliative tendency between the High and Low Fear groups ($p < .07$, $t = 1.96$) represents a replication of Schachter's (1959, p. 18) results. While the mean difference between High and Low Anxiety was even larger than that between the Fear conditions, it only approached significance ($p = .16$, $t = 1.46$) due to the marked heterogeniety of variance of the High Anxiety group.

TABLE 2
RELATIONSHIP OF MOTIVE TO SOCIAL AFFILIATION

	Mean Affiliation Strength[a]	Number of Subjects Choosing	
		Together	Alone or "0-Together"
Fear			
Low	34.0	12	3
High	51.0	19	1
Anxiety			
Low	27.0	11	4
High	8.0	10	12

Interaction: (Motive × Level) $p < .05$, $t = 2.30$, $df = 68$.

[a] The larger the score, the greater the affiliation tendency; isolation intensity score subtracted from affiliation intensity score.

Reasons Given for Affiliation Choice

The final measure taken was a questionnaire that explored the reasons the subjects gave for choosing to wait together with others or to wait alone. The 11 structured items on the questionnaire each presented a possible motive for affiliation; and each was accompanied by a 70-point scale on which the subject indicated how important he thought the motive was in determining his choice. The highly significant interaction between experimental treatment and questions ($p < .001$, $F = 3.74$, $df = 30.570$) on a repeated-measurement analysis of variance justified a search for those questions (motives for affiliation) that differentiated the groups.

Since there were too few subjects choosing the alone condition, the analysis is limited to those wanting to affiliate. The motives for affiliation that were most important for the High Fear subjects and most distinguished them from the Low Fear subjects were (the lower the mean, the greater the importance; 10 = extremely important):

1. I am not sure whether I am reacting in the same way as the others to the prospect of getting shocked and would like to compare my reactions to theirs. [Emotional comparison] High Fear $\bar{x} = 38$, Low Fear $\bar{x} = 54$, $p < .001$.

2. I feel worried about getting shocked and would like to know to what extent the others are worried too. [Extent of comparison] High Fear $\bar{x} = 40$, Low Fear $\bar{x} = 61$, $p < .001$.

3. I want to be distracted in order to take my mind off the prospect of getting shocked. [Distraction] High Fear $\bar{x} = 44$, Low Fear $\bar{x} = 59$, $p < .01$.

4. I am worried about the prospect of getting shocked and felt that talking with someone about it would get it off my chest. [Catharsis] High Fear $\bar{x} = 50$, Low Fear $\bar{x} = 59$, $p < .05$.

The reasons for affiliation given spontaneously to a single open-ended question also reflect the importance of these same considerations. Among High Anxiety subjects choosing to be alone, the major reason given spontaneously and supported by the scaled questions is the desire "to be alone to think about personal affairs and school work."

Curiosity as to "what the others were like" was important, but equally so across all conditions. Of least importance among all subjects are the following motives for affiliation ("oral stimulation" substituted for "shock" for Anxiety groups):

"It would be clearer in my own mind as to how I feel about getting shocked if I could express my reactions to someone else." "I anticipated that the others would offer reassuring comments." "I want to be with others to reassure myself that I am not the only one who was singled out to be shocked." "I feel that perhaps together we could somehow figure out a way to avoid getting shocked."

There are several large differences between the High Fear and High Anxiety groups; with the former finding the following motives as significantly more important: emotional comparison, extent of comparison, distraction, catharsis, and the physical presence of others ($p < .05$ in each instance). Similarly, an internal analysis of the High Fear group reveals these same motives (especially catharsis and emotional comparison) to be more important for those subjects who chose to affiliate most strongly than for those below the group median in affiliation strength.

Ordinal Position and Its Relation to Affiliation

While the reasoning used in the planning of the present study did not include predictions of the effects of ordinal position upon affiliation, data relevant to this question was nevertheless obtained, to check on Schachter's (1959) finding that affiliation tendencies increased with emotional arousal only among first- and only-born children. This finding is duplicated in the present study. First-born children want to affiiliate significantly more than later-borns under conditions of high fear, but not when the level of fear is low. While the mean affiliation intensity for the first-born High Fear subjects was 62, it was only 23 for the later-born High Fear subjects ($p = .05$, $t = 2.10$). This same general finding holds for the High Anxiety group, but again the within-group variability does not permit the large mean difference obtained (16 for first-borns and -3 for later-borns) to be statistically significant.

DISCUSSION

Since our basic hypothesis has been supported, our results lend credence to the previously drawn conceptual distinction between fear and anxiety. In view of the fact that our anxiety arousing stimulus was specifically designed to tap only one kind of repressed motive, it of course remains an empirical question whether or not the evocation of other types of presumably repressed motives also leads to social isolation.

In order to predict the consequences of the arousal of a motive, therefore, it is necessary to know which responses are required to reduce its tension. The probability of the social comparison response is, thus, a function of: the kind of motive aroused, the intensity of the motive, the degree of novelty of the emotional experience, the response hierarchy associated with the specific motive, and certain attributes of those with whom the person is to affiliate.

We do not question the assumption that the need for some kind of cognitive-emotional clarity and structure is a basic human motive. However, we feel that the need for self-evaluation is not the *most* salient motive aroused in the experimental situations that Schachter (1959) and we employed. We do not view the cognitive need to structure a vague emotional state as the primary motive in these experiments; we see social comparison not as an end in itself but merely as one of the several responses that are *instrumental* in reducing the tension associated with the situationally more salient motives of fear and anxiety.

Strict application of the theory of emotional comparison processes to the present experimental situation should lead one to predict greater affiliation tendencies for the High Anxiety subjects than the High Fear subjects, since the Anxiety situation was more unusual than that of Fear, and the emotion aroused was probably more novel and vague. The opposite

prediction, supported by the results, demands an approach, such as the one followed here, that specifies the probability of the response alternatives evoked by the dominant motives aroused.

As the emotional experience becomes very novel and unusual, the need for comparison of one's reactions with others should increase, and, hence, intensify affiliation tendencies. The induction of esoteric states of consciousness by "anxiety producing drugs" (being studied presently by Schachter) may be the kind of situation in which emotional comparison theory offers the best explanations and predictions. Under such circumstances, it may be possible to create emotional states that are epiphenomena of motives whose neurophysiological bases had never previously been set into motion. A more natural counterpart of this novel emotional experience occurs the first time a person experiences the emotions associated with the death of a loved one.

The predictive importance of knowing the specific responses appropriate to the motive aroused is clearly illustrated by the following examples. If a person's guilt is aroused, his response to feelings of guilt should be to seek out others only if they could be expected to punish him and, thus, to expiate his guilt, but not to affiliate with individuals perceived as unable to fill this role. Similarly, if repressed homosexual anxieties are aroused, isolation should generally be preferred to affiliation, as with oral anxiety in the present study. However, affiliation tendencies should increase if the subject is allowed to wait with females, but not if he can wait only in the company of males.

While our questionnaire data offer support for the importance of emotional comparison, they also point up the role of other motives such as need for catharsis and distraction. The marked difference in the importance of the reasons given for affiliation between the High Fear and High Anxiety groups is perhaps the most substantial evidence that the experimental manipulations have indeed led to the arousal of two quite different motives.

A final point of interest concerns the data about ordinal position. The finding that first-born children show greater affiliation tendencies than later-born children when either fear or anxiety are aroused supports Schachter's (1959) results. Theoretical and experimental attempts to uncover the dynamics underlying this "static" variable should prove interesting and fruitful.

SUMMARY

This experiment tests the utility of the psychoanalytic distinction between fear and anxiety for making differential predictions about social affiliation. It also assesses the breadth of generalization of Schachter's (1959) empirical finding of a positive relation between emotional arousal and affiliation. Seventy-two subjects were randomly assigned to four experimental treatments in which low and high levels of fear and anxiety were manipulated. The success of these inductions was established by a projective device and questionnaire data. The dependent variable of social affiliation was measured by having the subjects choose to await the anticipated exposure to the stimulus situation either alone or together with others.

The results show that, while the desire to affiliate increases as fear increases (a replication of Schachter's, 1959, results), the opposite is true for anxiety; as anxiety increases the desire to affiliate decreases. Thus, as predicted, our findings lend empirical support to the theoretical distinction between fear and anxiety. At the same time, our results suggest that the theory of social comparison processes may not be adequate to account for the general relationship between emotions and affiliative tendencies.

REFERENCES

FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117–140.

FREUD, S. *Inhibitions, symptoms, and anxiety.* (Originally published 1936) London: Hogarth, 1949. (a)

FREUD, S. *New introductory lectures on psychoanalysis.* (Originally published 1933) London: Hogarth, 1949. (b)

GERARD, H. B., & RABBIE, J. M. Fear and social comparison. Unpublished manuscript, Bell Telephone Research Laboratories, 1960.

SARNOFF, I. Psychoanalytic theory and social attitudes. *Publ. opin. Quart.*, 1960, **24**, 251–279. (a)

SARNOFF, I. Reaction formation and cynicism. *J. Pers.*, 1960, **28**, 129–143. (b)

SARNOFF, I., & CORWIN, S. M. Castration anxiety and the fear of death. *J. Pers.*, 1959, **27**, 374–385.

SCHACHTER, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

(Received March 16, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 364-366

MICROGENETIC ANALYSIS OF WORD ASSOCIATIONS IN SCHIZOPHRENIC AND BRAIN DAMAGED PATIENTS

NORMAN A. MILGRAM[1]

Boston University

In a recent study Flavell, Draguns, Feinberg, and Budin (1958) hypothesized that normals under time pressures and schizophrenics under the usual word association condition would give fewer microgenetically "mature" and more "immature" responses to stimulus words than normal controls. Such response categories as Synonym, Supraordinate, and Subordinate were assumed to reflect later microdevelopmental phases of cognition because of their logical "secondary process" character, while Distant, Completion, Perseveration, and other categories were assumed to represent earlier or abortive phases because of their superficial, contextual, or affect-mediated connection to the stimulus word. Differences between schizophrenics and normals were significant and in the predicted direction for 5 of the 15 categories scored, and it was concluded that so far as word association is concerned, microdevelopmental immaturity is accurately characterized by some but not all of the attributes assigned to it. The present study reports the results of microgenetic analysis of word associations in a second sample of schizophrenic patients and in a sample of brain damaged patients. This latter group also presumably gives expression to concrete or "immature" responses that normals usually suppress in favor of higher-level conceptual responses (Goldstein, 1943; Werner, 1948).

METHOD

The subjects were 32 schizophrenic patients and 30 brain damaged patients on the psychiatry and neurology wards of several VA hospitals and 32 normal controls. These male groups were equated on age, IQ (as measured by the Vocabulary subtest of the Wechsler-Bellevue Form I), and education. Means and *SD*s are reported in Table 1, none of the group differences approaching significance. Subjects were administered 50 words from the Kent-Rosanoff list with instructions to respond with the first word that came to mind. Tests were given individually and responses recorded by hand. The protocols were scored by the author in terms of Flavell's (Flavell et al., 1958, pp. 2–3) criteria as illustrated below:

A. Immature Categories

1. Completion: Completes a common word or phrase (whistle-"stop").

2. Distant: Unrelated or distantly related (music-"hospital").

3. Perseveration: Distantly related to its stimulus word, but obviously related to or identical with a stimulus or response word occurring earlier in the series.

4. Perseveration-meaningful: Closely related to its stimulus word, but identical to a stimulus or response word occurring earlier.

5. Clang: Related only by sound (earth-"worth").

6. Emotional: Connotes affective judgment (health-"wonderful").

7. Repetition: Repetition or grammatical variant (deep-"depth").

B. Mature Categories

1. Synonym: (chair-"seat").

2. Supraordinate: Denotes class of which stimulus word is member (butterfly-"insect").

3. Subordinate: Denotes member of class signified by stimulus word (sickness-"flu").

C. Categories of Indeterminate Microgenetic Phase[2]

1. Attribute: Modifies or is modified by stimulus word (beautiful-"daughter" or stove-"hot").

2. Verb: Response is verb and stimulus word a noun or vice versa (cheese-"eat" or eating-"food").

3. Contrast-Coordinate: Antonym or complementary (dark-"light" or salt-"pepper").

4. Multiword: (chair-"to sit on").

The word list[3] used contained nine words in common with Flavell's (Flavell et al., 1958) list and was composed of nouns, adjectives, verbs, and other grammatical classes in roughly the same proportions as his. Identical word lists would make the results of the two studies more comparable, but positive results with a different list would lend greater support to the microgenetic hypothesis. It was recognized that the frequencies of responses in the categories described above are partly a function of the parts of speech of the stimulus words. Nouns generally elicit other nouns, adjectives, and verbs; verbs rarely elicit other verbs or adjectives.[4]

[1] Now at the Nebraska Psychiatric Institute.

[2] Blocking or the failure to respond within 10 seconds was scored in Flavell's, but not in the present study.

[3] Word association list and instructions to the subjects may be obtained without charge from Norman A. Milgram; 602 South 44th Avenue; Omaha, Nebr., or from the American Documentation Institute. Order Document No. 6744 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilms or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

[4] Personal communication from Roger Brown.

In this important respect, the two lists were comparable.

RESULTS AND DISCUSSION

The microgenetic hypothesis that schizophrenic and brain damaged patients would respond with fewer "mature" and more "immature" associations than normals found little support. Of the significant group differences obtained, as many were in the direction opposite to prediction as in the predicted direction. Table 2 indicates that normals gave fewer Perseveration ($p < .05$) and Repetition ($p < .001$) than schizophrenics. On the other hand, they also gave fewer Supraordinate ($p < .001$) and more Perseveration-meaningful responses ($p < .01$). In addition, of the indeterminate categories that are considered concrete and possibly immature, normals gave more Verb and Attribute and fewer Multiword responses ($p < .001$ for each). The brain damaged group gave fewer Perseveration ($p < .05$), more Perseveration-meaningful ($p < .01$) and more Verb responses ($p < .01$) than schizophrenics, and more Multiword ($p < .001$) and fewer Attribute responses ($p < .01$) than normals.

In general, the brain damaged and schizophrenic groups were more similar to one another than was either to the normal controls. Rank order correlations of the frequencies of 13 response categories yielded coefficients of .66 and .77 for schizophrenics-normals and brain damaged-normals, respectively, and .95 for schizophrenics-brain damaged.

In retrospect, several of the underlying assumptions of the microgenetic approach to word association appear untenable. For one, it was assumed that the occurrence of a verbal response logically defined as mature was necessarily contingent upon the occurrence of an analogous mature cognitive process in the subject giving the response. Goldstein and Scheerer (1941, pp. 11–12), Yacorzynski (1941), and Rapaport (1945, p. 149) have illustrated this *non sequitur* and have contrasted apparent stability of conceptual word usage and actual deficits in conceptual thought processes in these patient groups. The word association task may be a poor choice for the illustration of microgenetic phases of thought because it permits

TABLE 1
AGE, IQ, AND EDUCATION OF THREE GROUPS

Group		Age in Years	Verbal IQ	Grade Completed
Schizophrenic	*M*	43.1	114.9	10.8
	SD	13.4	15.7	3.2
Brain damaged	*M*	45.6	110.5	11.5
	SD	15.2	17.7	3.2
Normal	*M*	44.5	115.8	11.6
	SD	15.2	10.3	3.0

TABLE 2
CATEGORY SCORES FOR THREE GROUPS

Category	Normal		Schizophrenic		Brain Damaged		p[a]
	M	*SD*	*M*	*SD*	*M*	*SD*	
Immature							
Completion	2.78	2.22	2.19	2.19	1.63	1.40	*ns*
Distant	1.09	1.35	2.06	1.87	1.43	1.75	*ns*
Perseveration	0.13	0.33	0.44	0.71	0.13	0.35	.05
Perseveration-meaningful	4.91	1.97	3.81	2.31	5.60	2.85	$<.05$
Clang[b]	0	0	0.07	0.24	0.03	0.17	*ns*
Emotional	1.28	1.25	0.88	1.45	1.23	1.96	*ns*
Repetition	0.13	0.33	0.97	1.26	0.43	0.90	$<.001$
Mature							
Synonym	6.00	3.72	7.84	3.93	7.47	4.50	*ns*
Supraordinate	1.06	1.14	3.38	3.10	2.53	2.12	$<.001$
Subordinate	0.81	0.78	0.81	1.09	0.93	0.93	*ns*
Indeterminate							
Attribute	4.09	2.69	2.09	2.39	2.63	1.67	$<.001$
Verb	4.19	2.18	2.31	2.15	3.90	2.37	$<.01$
Contrast-Coordinate	10.41	5.80	9.03	6.67	8.53	7.09	*ns*
Multiword	0.56	1.16	4.50	6.34	5.87	6.46	$<.001$

[a] All tests were two-tailed. F and t tests were used when the assumption of homogeneity of variance was accepted, χ^2 and t test with Snedecor's correction when it was rejected (Bartlett's test).
[b] The number of Clang responses was so few that the category was disregarded in subsequent comparisons.

subjects to fall back on well-practiced and overlearned verbal habits whose operations may circumvent the microgenetic route.

Flavell et al. (1958) suggested that the microgenetic formulation was hardly rigorous enough in its present form to rule out alternate explanations in terms of the inability of schizophrenics to maintain a constant mental set in word association. This writer agrees but would question if even normal subjects adopt the *same* mental set as one another, much less maintain a *constant* mental set in word association. In a recent study, Milgram (1960) showed that the set to give "adult" associations resulted in an increase in abstractly related responses and the reverse for "child" associations. These findings were obtained in a multiple-choice word association task which is hardly equivalent to the usual open-ended word association task. Nevertheless, these findings suggest that when the subjects are not instructed to adopt a specific mental set, some sophisticated subjects may choose to associate "immaturely" as evidence of their imaginative and creative approach, while others may adopt a "mature" semantic set as evidence of their knowing what words mean. This possibility receives some support from Flavell's own data on hospital aides and college students, who were significantly different on five response categories in the direction suggested above.

In summary, while the microgenetic approach toward cognitive processes in normal and pathological groups appears provocative and fruitful, attempts to apply this approach to word association are handicapped because of uncontrolled factors operating in the mental sets adopted in word association and because of a lack of necessary correspondence between the logical maturity of a verbal response and the actual maturity of the cognitive process which produced it.

REFERENCES

FLAVELL, J. H., DRAGUNS, J., FEINBERG, L. D., & BUDIN, W. A microgenetic approach to word association. *J. abnorm. soc. Psychol.*, 1958, **67**, 1-7.

GOLDSTEIN, K. The significance of psychological research in schizophrenia. *J. nerv. ment. Dis.*, 1943, **97**, 261-279.

GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior: An experimental study with special tests. *Psychol. Monogr.*, 1941, **53**(2, Whole No. 239).

MILGRAM, N. A. Cognitive and empathic factors in role-taking by schizophrenic and brain-damaged patients. *J. abnorm. soc. Psychol.*, 1960, **60**, 219-224.

RAPAPORT, D. *Diagnostic psychological testing*. Vol. 1. Chicago: Year Book Publishers, 1945.

WERNER, H. *Comparative psychology of mental development*. Chicago: Follett, 1948.

YACORZYNSKI, G. K. An evaluation of the postulates underlying the Babcock Deterioration Test. *Psychol. Rev.*, 1941, **48**, 261-267.

(Received March 17, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 367-374

A FREE RESPONSE APPROACH TO THE STUDY OF PERSON COGNITION[1]

LEEROY BEACH AND MICHAEL WERTHEIMER[2]
University of Colorado

In experiments in "person perception," *S* (the subject) is ordinarily asked to make judgments about a particular person (hereafter called *O*) in a given situation. Typically the *E* (experimenter) requires *S* to use fixed scales for specifying his judgments of the particular *O* who is the object of the perception or cognition. Often *E* probably requires *S* to make judgments on dimensions that *S* would not ordinarily use in "sizing up" a particular *O* in that situation. Interpretation of the results of such studies becomes difficult because the "meaning" and "salience" of the scales for *S* may be unclear, atypical, or even absent. Although the fixed scale technique has the advantage of easy calculations, a technique is also needed by which *E* could determine which scales are appropriate for the *S*s, *O*s, and situations of interest to him. Interpretation of experimental results could be facilitated by using only those scales, in a fixed scale study, which were previously shown to be meaningful for the *S*s, the *O*s, and the situation under consideration.

In their discussion of the necessity for attempting to determine the relevant dimensions on which *S*s make judgments of *O*s, Hastorf, Richardson, and Dornbusch (1958) "... urge a momentary pause in the search for the 'trait' of sensitivity in order to study the qualities of a person's experience of others in terms of the verbal categories he uses in reporting that experience. The central characteristic of this type of methodology would be the eliciting of free and unrestricted descriptions of other persons. The classes of persons to be described would be specified by the experimenter" (p. 56). One purpose of this paper is to demonstrate the feasibility of such a methodology.

In addition to the questionable meaningfulness of scales and other response measures used in research on person perception, the stimulus used for the *O* is often highly unrepresentative of the population to which results are to be generalized. The literature shows that the *O* who is to be the object of the perception or cognition on the part of *S* has been specified in many ways. One can use real people (e.g., Newcomb, 1956), films (e.g., Stone, Leavitt, & Gage, 1954), recordings (e.g., Jones & deCharms, 1958), handwriting (e.g., Huntley, 1940), photographs (e.g., Cline, 1956), drawings of faces (e.g., Brunswik, 1956), stories (e.g., Cawthorne & Kogan, 1959), autobiographies (e.g., Sperber, 1959), lists of adjectives (e.g., Asch, 1946), abstract figures (e.g., Heider & Simmel, 1944), etc. Many of these studies use artificial or synthetic *O*s and attempt to generalize to the perception of real, living people. While in many cases the stimuli serve the purpose of the research in which they were used, their generalizability is somewhat tenuous. *S*s are doubtless aware of the artificiality of the situation and it is not improbable that these conditions also influence the results obtained.

When asking *S* to judge real people *E* has difficulty in knowing, manipulating, and controlling all of the variables that may be relevant from *S*'s point of view. Not only is it difficult to obtain a broad enough sample of *O*s to allow generalization (often the sample is restricted to classmates or fellow dormitory residents and the like) but *E* often does not know, for instance, how well the *S* and *O* know each other, how long they have been friends, *O*'s perceived status from *S*'s point of view, etc.

S moves, daily, in a world inhabited by different persons having certain relationships to him; some are of the same sex, some of the opposite, some are well known to *S*, others are only undifferentiated figures with no clear

[1] This study was supported in part by the Behavioral Science Division of the Air Force Office of Scientific Research, under contract with Georgetown University, and in part by the University of Colorado's Council on Research and Creative Work.

[2] We would like to thank Glen Brennan for his assistance in some of the data analysis, and O. J. Harvey, William A. Scott, and Aldora Graf for their helpful comments and suggestions in connection with this report.

identity, some are very important to *S*, others of little concern to him, etc. If one were to select a number of these people one would probably find that *S* describes them differently and behaves toward them differently. Further, the requirements of the situation doubtless influence what dimensions or characteristics assume the greatest importance in *S*'s descriptions of these people.

The *general* research question becomes, what kinds of dimensions are used by different kinds of *S*s when describing different kinds of *O*s under different kinds of conditions?

Kelly (1955) has provided the Role Construct Repertory Test as a method of investigating this question. This test was designed for individual administration; when it is group administered it is difficult for *E* to decide if two *S*s mean the same thing by a given term or expression; a term needs to be embedded in a context to be adequately defined. In the individual administration of the test *E* can get additional information from *S* that would aid in this decision. Further, the Role Construct Repertory Test assumes that all dimensions apply, in one way or another, to all *O*s described and that all dimensions are bipolar (Kelly, 1955, p. 268), assumptions that the present *E*s feel may be premature.

The present study attempts to apply another method to provide information about the kinds of dimensions used by different kinds of *S*s when describing different kinds of *O*s. The method is simply to sample the people from *S*'s world and obtain his impressions of them. But the *O*s sampled from different *S*s' worlds should be comparable so that some degree of generalizability can be attained. To guarantee that *S*s make their judgments about real people, that these *O*s represent a rather broad selection of persons, and at the same time to assure some degree of comparability of *O*s over various *S*s, a method somewhat similar to that used by Kelly's Role Construct Repertory Test was adopted, in that specific classes of *O*s to be described were designated. *S* was asked to choose someone he knew who fit the class, and then to describe the person he chose using any terms or dimensions he wished. After this a content analysis[3] was performed on the obtained free response data in an attempt to discover relationships. These relationships may serve to support the feasibility of using this method for assessing the dimensions used by the *S*s and to show that dimension use relates to differences in the kinds of *S*s doing the describing and differences in the kinds of *O*s being described.

The free response—content analysis approach involves the following assumptions: *S*s can give verbal descriptions of other people. These descriptions employ dimensions that *S* feels are relevant in the judging of *O*. (It may be argued that *S* may, for psychodynamic reasons, be unable to verbalize on the most important dimensions. But as a beginning, the present procedure may prove meaningful, at least for conscious judgments.) The relative frequency of use of different dimensions by a given *S* is a reflection of the relative salience or importance of the dimensions for that *S*.

Such considerations lead to expectations like the following:

1. The content of free verbal descriptions of people can be readily and reliably analyzed. This implies that a content analysis can be performed, and that all responses can be reliably coded into a set of selected categories. Some aspects of the adequacy of the category system used can be assessed by the usual criteria of reliability such as interjudge agreement or repeat scoring, and to a degree, by the extent to which meaningful and statistically significant results can be obtained when the category system is used in evaluating various hypotheses. A related expectation is that a given *S*'s use of categories for a particular kind of *O* and over various types of *O*s should be consistent over time.

2. *S*s "weight" dimensions differently when describing different *O*s; category use varies with the kind of *O* being described. That is, *O*s different in certain respects tend to be described with emphasis on different cate-

[3] There is, of course, no reason to believe that the free response—content analysis technique reveals the "true" dimensions uncontaminated by the content analyst; it does, however, help one gain a better perspective on the variety of dimensions used "naturally." The investigator must make the decisions about what will actually be used in a fixed scale study. It is hoped that the perspective gained from the proposed approach will prevent him from demanding that the *S*s use dimensions which, from their point of view, are inappropriate or irrelevant.

gories, with different evaluative tone, and with different amounts of information.

3. Similarly, groups of *S*s that differ in certain respects tend to differ in their "weighting" of categories when describing the same kinds of *O*s. Here too there should be differences in evaluative tone, and different amounts of information. Expectations 2 and 3 state that descriptions vary as a function of *S* differences, when certain *O* characteristics are held constant, and with certain *O* differences, when certain *S* characteristics are held constant.

4. One might also expect an interactive effect between the two sets of variables, such that the same *O* differences produce different differences in the descriptions when different groups of *S*s are used.

METHOD

Subjects

Sixty-six students (30 female, 36 male) enrolled during the 1958 summer session in an introductory psychology course at the University of Colorado served as *S*s.

Object Persons

As already indicated, the *O*s were designated by constructing classes of people and having the *S*s think of persons who fit the class. The classes were formed by using all possible combinations of well known–not well known, higher–equal–lower status (relative to *S*), and same and opposite sex, to generate 12 classes of *O*s to be described. An example of a class designation used is: "Think of a person who is the same sex as you, whom you know well, and whom you consider of higher status than yourself."

Procedure

During a regular 60-minute class meeting the *S*s were given 8½″ × 5½″ booklets, the first page of which contained instructions stating that *S* would be asked to think of various people he knew and to give descriptions of them.

At the top of each of the 12 following pages there was a short class designation like the one given above. Immediately below this *S* was asked to indicate *O*'s age in a space provided for this purpose; the rest of the page was to be used for *S*'s description of *O* ("Describe this person"). The age question was included to help assure that *S* actually thought of a real *O*. If *S* could not think of a person to fit the designation he was requested to leave the page blank.

In addition to the 12 descriptions of *O*s, *S* was also asked to describe himself. A page for this self-description occurred first in the booklet for 24 *S*s (12 males, 12 females), in the middle of the booklet for 18 (12 males, 6 females), and at the end for 24 (12 males, 12 females). In order to minimize possible systematic order effects, all but the instructions and final page (which asked for information such as *S*'s age, his major, etc.) in the book were arranged in different orders in different booklets, with no more than two *S*s (1 male and 1 female) having the same page order.

Content Analysis

To arrive at content categories for the analysis, a judge (Judge 1) went through the booklets to obtain a global impression of what dimensions were used. This procedure yielded a sizable list which was then grouped on the basis of similarity into 13 major categories.[4] These categories and their component dimensions are presented in Table 1.

The categories may be viewed as falling into four major divisions:

I. Objective Information. *O*'s appearance (Category 1), his background (Category 2), and his job, income, etc. (Category 3).

II. Social Interaction. *O*'s behavior toward *S* (Category 4) and his behavior toward other people (Category 5). *S*'s reaction to and acceptance of *O* (Category 7) and other people's reaction to and acceptance of *O* (Category 6).

III. Behavioral Consistencies. *O*'s temperament (Category 8), self-concept and emotional adjustment (Category 9), and his morals and values (Category 10).

IV. Performance and Activities. *O*'s abilities (Category 11), motivation (Category 12), and his activities and interests (Category 13).

Using the 13 content categories, Judge 1 scored all of the booklets, without reference to the designations of the *O*s at the top of each page. Scoring consisted of recording which content categories were used, and the frequency of use for each page of each booklet.

Three procedures were used to check scoring reliability. Two other judges (Judges 2 and 3) were asked to perform the analysis independently on 10 of the booklets. Judge 1 reanalyzed 10 booklets 6 months after his first scoring. Eleven *S*s returned 2 weeks after the initial study to fill out the booklets again. At the time of this retest the *S*s were told to make no special effort to describe the same *O*s they had described previously, nor to make an effort to think of new people. Both the test and the retest protocols were analyzed by Judge 1.

For purposes of data analysis, a summary data sheet was constructed for each booklet with the 13 content analysis categories across the top and the 12 *O*s and the self down the side. This gave a matrix of 169 cells, the rows of which were the descriptions of *O*s and the columns of which were the content categories.

The judge, when reading a given description, placed a mark to indicate the presence of a category, in the appropriate column cell for each phrase or complete

[4] Obviously other judges might have arrived at different categories, but in light of the results to be reported the ones used here do appear to be meaningful. The investigator using a free response approach preparatory to a fixed scale study must, of course, try as best he can to arrive at categories which are suitable to his particular research.

TABLE 1

CONTENT CATEGORIES AND EXAMPLES OF DIMENSIONS CONTAINED IN THEM

Category	Dimensions
1. Appearance of *O*.	References to build, features, dress, general good or bad looks, sex appeal, voice quality.
2. *O*'s background.	Family solidity, parents' occupation, family status and wealth, nationality, personal childhood experiences (e.g., "He was overshadowed by his sister").
3. General information.	*O*'s present marital status, education, occupation, church preference, veteran, college, college year, where and how *S* met *O* if this is "merely information."
4. *O*'s behavior toward *S*.	Specific mention of some behavior on the part of *O* toward *S* (e.g., "He has always been nice to me").
5. *O*'s behavior toward other people.	Specific behavioral modes involving other people (which may or may not include *S*). Conformity, behavior reflecting some sort of general attitude (snooty, overbearing, etc.), degree of openness in interactions, willingness to be with people and to interact with them in the usual socially acceptable manner. This does not include judgments about *O*'s "nature" (e.g., "He has a *friendly* (not *good*) personality"). Examples are: charming hostess, poor manners, likes people, easy to get to know, etc.
6. Other's reaction to *O*.	Opinions of *O* held by other people and their acceptance or nonacceptance of him. Such things as: people usually like him, he isn't very popular, etc.
7. *S*'s reaction to *O*.	General or global pronouncements about *O*. Such statements as: he has a good personality, I like him very much, the best friend I ever had, likeable, enjoyable to be around, easy to get along with, etc. This category also includes any guesses *S* makes about *O*'s future or about how things are going for him at present (e.g., he'll never reach his goal, his life seems pleasant, etc.).
8. *O*'s enduring personal characteristics.	Rather consistent personality factors which dictate the general color of his interactions with other people without actually being an instance of that behavior. Examples are: quiet type, has a sense of humor, extrovert, outgoing, practical, conservative, gushy, I-don't-care-attitude, sincere, serious, etc.
9. *O*'s emotional adjustment and self-concept.	Emotional adjustment and *O*'s behavior toward himself from *S*'s point of view, not necessarily from a psychologist's. Examples are: He's confused, she feels sorry for herself, inferiority complex, mature, knows when to work and when to play, humble, wonderful outlook on life, spoiled, etc.
10. *O*'s values and ideals.	Any set values or standards which *O* imposes or fails to impose upon himself or generally holds as desirable in people. Examples are: religiosity, morals, political values, mercenary, drinking and smoking habits, etc.
11. *O*'s abilities.	*O*'s intellectual abilities, other abilities (mechanical, etc.), talents, scope of knowledge, proficiency, etc. Also achievements which reflect ability. This category does *not* include social abilities.
12. *O*'s aspirations, ambition, and motivation.	Goals in life, direction in activity, motivation in tasks undertaken, etc.
13. *O*'s interests.	Interests aside from specific goals (occupation or studies). Not social unless a specific interest in working with or being with people (such as camp work). Also includes offices held and memberships in other than purely scholastic organizations. Examples are: likes sports, out for a good time, likes to have fun, belongs to social groups, has many hobbies, etc.

thought in the description. A plus sign was used if *E* considered the phrase or thought to suggest a positive or favorable evaluation of *O* by *S*, a zero if it is was neutral, and a minus sign if it was negative or unfavorable. Number of units was then summed across rows and down columns. The row and column sums (and, for certain analyses, the cell sums), were converted into proportions of the overall total, to correct for individual differences in "wordiness" or total number of "units." According to our assumptions, the proportion of total responses in each *column* provides an index of the relative weights of the categories. The proportion of total responses in each *row* gives the relative response productivity for each *O* described (or for the self-description). Comparable proportional transformations were undertaken for the evaluative symbols.

These proportions make it possible to analyze the effects of the independent variables on category use. For example, to study how degree of acquaintance affects category use, one need only compare the average column proportion for well-known *O*s with that for not well-known *O*s for each *S* over the whole group of *S*s. Or, to study response productivity as a function of degree of acquaintance, one compares average row proportions to see whether more was said about well-known than about not well-known *O*s. Such analyses require the use of some statistical test such as *t* or the Wilcoxon signed ranks tests. Nonparametric tests were

used throughout the analysis to avoid assumptions regarding variance and distribution.

RESULTS

The following results provide information about each of the expectations stated above: *The content of free verbal descriptions of people can be reliably analyzed.* Reliability was assessed by interjudge agreement and by rescoring after a time interval; the retest results are also relevant to the reliability question.

For the reliability analysis, agreement between two scorings of the same *S*'s protocol was studied by cell-for-cell comparison of the two summary data sheets for the protocols. Fourfold tables were constructed, using the following cell entries: both judges agreed on the presence of the category in the description, Judge A said it was present but Judge B said it was not, Judge B said it was present but Judge A said it was not, and both judges agreed on the absence of the category. Since *N*s were small (about 10 protocols for each reliability check), the fourfold tables were combined across *O*s for each of the 13 categories. The phi coefficients were computed on these composite fourfold tables to provide a quantitative index of agreement. Table 2 presents the results for each category and for all categories combined, for each reliability check.

The table indicates that, if inter- and intrajudge agreements are considered, scoring reliability may well be considered satisfactory. Scoring of Categories 1, 2, 3, and 6, especially, seems reasonably reliable; scoring reliability of Categories 4, 7, 8, and 10 appears somewhat less satisfactory. The retest procedure in general yielded far lower agreement. In the last rows of the table, a simple mean phi for each category over all reliability procedures (by Fisher z transformation) appears for purposes of indicating relative reliability of the different categories in general.

The expectation that a given *S*'s weighting of categories for a particular kind of *O* and over various types of *O*s should be consistent over time was not clearly supported. As Table 2 shows, although only 1 of the 13 phi's was negative ($p = .01$) the mean phi coefficient over the 11 test-retest protocols was only .16. This coefficient is not significantly greater than zero, and indicates that there was no appreciable degree of consistency from the first protocol to the retest. Various considerations may clarify this result. First, most *S*s, when asked at the end of the retest session, reported that they had described some new *O*s in addition to some of the *O*s described during the first session. Examination of the retest protocols indicated that an average of 54% of the *O*s had not been described before. Second, such a test-retest procedure combines the error variance contributed by the judge with that contributed by the *S* and might therefore be expected to reduce the obtained phi coefficient. Third, the relative reliabilities of the categories for test-retest were comparable to those in the assessments of judge reliability. A Spearman rho of .61 ($p = .03$) was obtained between the mean interjudge reliability for each category and the test-retest reliability for each category. The relative reliabilities of the categories thus appear to be similar in the two checks.

It should be understood that this study was

TABLE 2

PHI COEFFICIENTS OBTAINED IN THE RELIABILITY ANALYSES

Comparison	Categories													Mean phi
	1	2	3	4	5	6	7	8	9	10	11	12	13	
Judge 1: Judge 2	.78	.69	.67	.49	.59	.70	.53	.39	.61	.48	.45	.52	.37	.56
Judge 1: Judge 3	.95	.79	.81	.00	.63	.74	.50	.45	.62	.44	.75	.63	.79	.62
Judge 2: Judge 3	.89	.83	.70	.27	.41	.57	.50	.44	.62	.32	.64	.63	.61	.57
Judge 1: Judge 1 (6 months later)	.94	.83	.86	.57	.85	.83	.60	.67	.77	.65	.75	.79	.62	.75
Test-Retest	.47	.09	.48	.00	.18	.03	.02	.13	.29	.07	.17	.15	−.05	.16
Mean phi (without retest)	.91	.79	.77	.34	.65	.72	.54	.49	.66	.48	.66	.65	.62	.66
Mean phi (with retest)	.86	.71	.72	.28	.58	.63	.45	.43	.60	.40	.58	.58	.52	.58

exploratory and that the results to be reported in the following sections are illustrative rather than exhaustive. They serve to show that data do support each of the expectations outlined in the introduction. (In the nonparametric statistical tests reported here the rejection criterion was .05, two-tailed, unless otherwise stated.)

Os different in certain respects tend to be described with different categories, different evaluative tone, and different amounts of information. All three aspects of this expectation were supported by a variety of results.

1. *Category Use.* Example: The proportion of total responses made for well-known *O*s is higher in Category 1 (physical characteristics) than for *O*s who are not well known.

2. *Evaluative Tone.* Example: There was a significantly greater proportion of negative evaluations of *O*'s motivation (Category 12) for lower status *O*s than for equal status *O*s, and for equal status *O*s than for higher status *O*s.

3. *Amount of Information.* Example: The amount of information provided in a description of an *O* varied with the degree of acquaintance. The largest proportion of units was used in self-description, a moderate proportion in describing well-known *O*s, and the smallest proportion when describing not well-known *O*s. Seventy-seven percent of the *S*s exhibited this order of response productivity.

These examples indicate that the selection of different kinds of *O*s to be described does seem to produce differences in the descriptions.

Ss different in certain respects use different categories, different evaluative tone, and different amounts of information. Once again, a variety of findings supported this expectation.

1. *Category Use.* Examples: Regardless of whether a male or a female *O* was being described, male *S*s tended to use categories that emphasized physical appearance, background, occupation, values, abilities, motivation, and interests (Categories 1, 2, 3, 10, 11, 12, and 13) more than did female *S*s. On the other hand, female *S*s tended to use content categories emphasizing social skills and adjustment (Categories 5 and 9) more than did men.

S's age was also related to category use. For example, Category 3 (objective information) was used more by male *S*s whose age was above the median than by those male *S*s whose age was below the median. (Male ages ranged from 18–29, with a mean of 21.94 and a standard deviation of 2.94.)

2. *Evaluative Tone.* Examples: Females were more likely to use evaluative terms than were males, in describing physical characteristics of the *O* (Category 1). Increase in age was associated with decrease in negative evaluations in Category 6 (other people's reaction to *O*).

3. *Amount of Information.* Example: Under the conditions of the study, female *S*s tended to be somewhat more productive than male *S*s; the number of codable units in the entire protocol indicated a greater amount of information in descriptions made by females than in descriptions made by males. (This was significant only at the .05 level, by one-tailed test.)

Descriptions vary as a function of the interaction of S and O characteristics. Such interactions were found to affect all three dependent variables, category use, evaluative tone, and amount of information.

1. *Category Use.* Examples: Category use followed a sex-sex, a sex-acquaintance, and a sex-status interaction. Thus male *S*s made greater use of Category 11 (abilities) when describing male *O*s than when describing female *O*s; on the other hand they made greater use of Category 7 (*O*'s behavior toward *S*) when describing female *O*s than when describing male *O*s. Neither of these findings held for female *S*s.

Male *S*s used Category 12 (motivation) more in describing well-known *O*s than in describing *O*s they did not know well, while female *S*s used this category equally for both kinds of *O*s. Female *S*s, on the other hand, used Category 13 (interests) more for well-known *O*s than for not well-known *O*s; male *S*s did not show this tendency.

Male and female *S*s showed a relation between use of Category 12 (motivation) and status of the *O*. Use of this category was greater for both equal and lower status *O*s than for higher status *O*s. An additional interactive pattern is that female *S*s used Category 6 (other people's reaction to *O*) less for equal and lower status *O*s than for higher, while male *S*s used it less for lower status and equally for higher and equal status *O*s.

2. *Evaluative Tone.* Example: A significant interaction was found between degree of

acquaintance with *O* and the evaluative tone of responses in Category 10 (values), as a function of *S*'s sex. For male *S*s an increase in degree of acquaintance was associated with an increase in positive evaluations; that is, well-known *O*s were more likely to be given positive evaluations in regard to values than were *O*s who were not well-known. Descriptions by female *S*s did not evidence this trend.

3. *Amount of Information.* Example: Male *S*s used a significantly greater proportion of units when describing male *O*s than when describing female *O*s, while female *S*s did not differ in response productivity as a function of *O*'s sex.

These findings illustrate the interaction effects of the kind of *S*s doing the describing and the kind of *O*s being described.

DISCUSSION

In the preceding examples we have presented only some of the more striking instances of results to illustrate the variety of approaches that can be taken with this sort of strategy.[5]

Although it would be desirable to extend the application of free response techniques to a wider range of *S*s, *O*s, and situations, these preliminary results indicate that the dimensions along which *S*s make judgments of *O*s are indeed dependent upon the characteristics of the *S*s, of the *O*s, and of the interaction between *S* and *O* variables. The present study suggests that in one situation it is possible to categorize *S*s' descriptions of *O*s adequately enough so that the differences between *S*s, the differences between *O*s, and the effects of various interactions between *S* and *O* variables can be shown to produce significant differences in category use, evaluative tone of the descriptions, and the amount of information given in descriptions.

These results suggest that when selecting scales or dimensions upon which *S* is to be required to make judgments about *O*s, one must keep in mind that these dimensions are not going to be applicable to all *O*s and *S*s. *S*, it would seem, considers different dimensions appropriate for different kinds of *O*s. How one is to interpret responses made on "inappropriate" dimensions is quite unclear. It is not impossible, for instance, that on these dimensions responses are made randomly. Or, perhaps responses on these inappropriate dimensions are based on stereotypes while responses made on dimensions that *S* considers appropriate are more specific to the *O* being judged. This might help to explain some of the results in the person perception literature (e.g., Cline & Richards, 1960) which suggest that accuracy in judging the stereotype accounts for a good part of overall accuracy in making judgments about a given *O*. In fact, in an accuracy experiment it may be necessary to find those dimensions which *S* considers appropriate to *O* and those dimensions which *O* considers appropriate to himself. It may be that only dimensions that are jointly considered appropriate yield a high degree of accuracy.

In general, a failure to obtain the predicted experimental results in studies of person perception or person cognition could be a result of incorrect hypotheses *or* of the inappropriateness of the dimensions selected for *S* to use in making responses concerning *O*. Consequently, it seems reasonable to attempt to find the dimensions which *S* considers relevant or appropriate to a given *O* or kind of *O* and then use these dimensions as the scales in the subsequent fixed scale study.

SUMMARY

In person perception experiments subjects (*S*s) have often been required to make judgments about other people (*O*s) on dimensions that may not be applicable or relevant from the *S*'s point of view. The results of such studies are, consequently, difficult to interpret. It has been argued that fixed scale research should be preceded by free response—content analysis studies to aid in decisions as to which scales or dimensions are appropriate for the *S*s, the *O*s, and the situation involved.

In an attempt to show that dimension use does differ for different *S*s and *O*s, *S*s were asked to describe persons they knew who fit the specifications of classes of *O*s designated by *E*. These descriptions were then content analyzed.

The findings serve to support the contention that the dimensions used in judgments of other

[5] Although the examples chosen here were of course selected from among a variety of analyses, it is unlikely that the differences reported were due to chance alone. Over one-third of all of the analyses (over 100) attempted on the data yielded differences significant at the .05 level or better, two-tailed.

people are a function of the *S*s doing the judging, the *O*s being judged, the situation, and the interaction among these variables.

It was concluded that a prior free response—content analysis approach like the present one might well be considered logically and methodologically prior to the study of how independent variables affect ratings along fixed predetermined scales.

REFERENCES

ASCH, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258–290.

BRUNSWIK, E. *Perception and the representative design of experiments*. Berkeley: Univer. California Press, 1956.

CAWTHORNE, FLORENCE M., & KOGAN, N. The differential cue value of age and occupation in impression formation. Paper read at EPA meeting, 1959.

CLINE, M. G. The influence of social context on the perception of faces. *J. Pers.*, 1956, **25**, 142–158.

CLINE, V. B., & RICHARDS, J. M., JR. Accuracy of interpersonal perception—A general trait? *J. abnorm. soc. Psychol.*, 1960, **60**, 1–7.

HASTORF, A. H., RICHARDSON, S. A., & DORNBUSCH, S. M. The problem of relevance in the study of person perception. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 54–62.

HEIDER, F., & SIMMEL, MARIANNE. An experimental study of apparent behavior. *Amer. J. Psychol.*, 1944, **57**, 243–259.

HUNTLEY, C. W. Judgments of self based upon records of expressive behavior. *J. abnorm. soc. Psychol.*, 1940, **35**, 398–427.

JONES, E. E., & DECHARMS, R. The organizing function of interaction roles in person perception. *J. abnorm. soc. Psychol.*, 1958, **57**, 155–164.

KELLY, G. A. *The psychology of personal constructs*. Vol. 1. New York: Norton, 1955.

NEWCOMB, T. M. The prediction of interpersonal attraction. *Amer. Psychologist*, 1956, **11**, 575–586.

SPERBER, Z. Rigidity of judges and their utilization of autobiographical material in making predictions. Paper read at EPA meeting, 1959.

STONE, G. C., LEAVITT, G. S., & GAGE, H. L. *Generality of accuracy in perceiving standard persons*. Urbana: Bureau of Educ. Res., Univer. Illinois, 1954.

(Received March 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 375-385

ADDRESS IN AMERICAN ENGLISH[1]

ROGER BROWN AND MARGUERITE FORD
Massachusetts Institute of Technology

WHEN one person speaks to another, the selection of certain linguistic forms is governed by the relation between the speaker and his addressee. The principal option of address in American English is the choice between use of the first name (hereafter abbreviated to FN) and use of a title with the last name (TLN). These linguistic forms follow a rule that is truly relational. Their use is not predictable from properties of the addressee alone and not predictable from properties of the speaker alone but only from properties of the dyad. Kinship terms of address (e.g., dad, mom, son) are also relational language, but they constitute a restricted language of relationship since most dyads that might be created in America would not call for any sort of kinship term. Proper names, on the other hand, constitute a nearly universal language of relationship; the semantic dimensions involved serve to relate to one another all of the members of the society.

Much of our knowledge of the dimensions that structure interaction has been obtained from paper-and-pencil performances such as the status ranking of acquaintances or the sociometric choice from among acquaintances. There can be no doubt that the dimensions revealed by such data are genuinely functional in ordinary life. However, performances elicited by the social scientist are not the everyday performances that reveal dimensions of social structure and cause each new generation to internalize such dimensions. It is always possible that a paper-and-pencil task imposes a dimension or encourages an unrepresentative consistency and delicacy of differentiation. It is desirable to study social structure in everyday life, but much of the everyday behavior that is governed by social dimensions is difficult to record and involves an uncertain number of significant contrasts (e.g., influencing decisions, smiling at someone, warmth of manner). Forms of address are speech and so there is a recording technique called writing which preserves most of the significant detail. English forms of address are reasonably well described by a single binary contrast: FN or TLN. These forms are ubiquitous and for any set of interacting persons it is possible to obtain a collection of address dyads from which the latent structuring dimensions can be inferred. This paper provides an example of the application of semantic analysis to the study of social structure. We begin by describing the norms of American address, then point out a certain pattern in these norms which is to be found in the forms of address of all languages known to us. This abstract speech pattern suggests a feature of social structure that may possibly be universal.

[1] This study was supported by a research grant (M-2649) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service.

METHOD

Materials

To discover the norms of address in American English we require a large sample of usage. The range of the subject population is vast but the uniformity must be great. Some sensible compromise is required between the stratified national sample dictated by the scope of the problem and the unsystematic observation of one's friends dictated by the probable simplicity of its solution. Four kinds of data have been used.

Usage in modern American plays. There are more instances of address in plays than in any other form of literature. Thirty-eight plays written by American authors, performed since 1939, and anthologized in three volumes of *Best American Plays* (Gassner, 1947, 1952, 1958) were used. Of the plays in these volumes we omitted only those set in some remote time or in a country other than the United States. Without listing all the titles it is possible to indicate the range of American usage that it represented. By geographic region we move from *The Philadelphia Story* to the Southland of *The Member of the Wedding* through the midwest of *Picnic* to the Far West of *The Time of Your Life*; by social class from *The Rose Tattoo* and *A View from the Bridge* through *Death of a Salesman* and *Tea and Sympathy* to *The Solid Gold Cadillac*. For military usage there are such plays as *Mister Roberts* and *No Time for Sergeants*. The speech of Jewish, Italian, and Negro minorities is represented and also that of such groups as politicians, corporation executives, narcotics addicts, college professors, and policemen.

Of course these materials are not a record of actual speech from the characters named but are the speech constructed for such characters by playwrights. Probably playwrights accurately reproduce the true norms

of address, and it is possible to check one author against another. In addition, however, there are three other kinds of data to catch possible inaccuracies in this first set.

Actual usage in a Boston business firm. For 2 months a man employed in a drafting firm took advantage of leisure moments to jot down for us instances of linguistic address overheard from his fellow workers. He collected address terms for 214 different dyads in which 82 different people are involved either as speaker or addressee. Each person is identified by full name, sex, age, and position in a 12-level occupational hierarchy. In addition, we benefited from our informant's good knowledge of friendship patterns among the dramatis personae.

Reported usage of business executives. Each year there is at MIT a group of Alfred P. Sloan Fellows; these are business executives between 30 and 38 years of age, who are nominated by their respective employers to study for one year at the institute. The 34 Sloan Fellows designated for 1958–59 served as informants for this study. They had come from many different parts of the country, most of them from very large corporations but a few from small companies.

By the time the Sloan Fellows served as informants the general pattern of the norms had become sufficiently clear to make possible the writing of a questionnaire designed to elicit the most important information. Each man was asked to write down the full names and positions of four persons whom he was accustomed to see nearly every day at his place of business, and he was to distribute his selections so as to include: one person equal to himself in the organization hierarchy with whom he was on close or intimate terms, one person equal to himself with whom he was on distant or formal terms, one person superior to himself in the organization hierarchy, one person subordinate to himself in the hierarchy. After listing the names the informant was asked to write down for each person listed the exact words that he (the informant) would customarily speak in greeting that person for the first time each day. In addition, the Sloan Fellows were asked some other questions that will be described when they become relevant.

Recorded usage in Midwest. The Psychological Field Station of the University of Kansas directed by Roger G. Barker and Herbert F. Wright (1954) kindly allowed us to make an extended study of 10 "specimen records," each of which records the events and conversation in a full-day of the life of a child. The station has also allowed us to work with a set of brief "behavior settings observations" made on 56 children in the town called Midwest and on 56 children in Yoredale, England, matched with those in Midwest by sex and age. In these materials we have been primarily concerned with a grammatical analysis of the kinds of "mands" addressed to children (for another study of the language of social relationship) but have, in addition, taken the opportunity to check our conclusions about American address against the Midwest records.

Procedure

For each of the 38 plays every instance of address was recorded, together with an identification of the speaker and addressee. The method of study is a sort of controlled induction.[2] Approximately one-third of the plays were first examined in an effort to discover rules that would summarize all of the instances of address they contained. The resulting provisional rules were then tested against a second set of plays and underwent some revision. The revised rules proved adequate to the description of all instances of address in a third and final set of plays. The supplementary data from a business firm, from the Sloan Fellows, and from the Midwest records were used as additional checks on the rules induced from the plays and also to test several particular hypotheses.

MAJOR PATTERNS OF ADDRESS

If we consider only the FN and TLN there are just three logically possible dyadic patterns: the reciprocal exchange of FN, the reciprocal exchange of TLN, and the nonreciprocal pattern in which one person uses FN and the other TLN. All three patterns occur with high frequency and the problem is one of inducing the semantic factors invariably associated with a given pattern and serving to distinguish it from the others.

In classifying instances of address into the three classes, FN was taken to include full first names (e.g., Robert), familiar abbreviations (e.g., Bob), and diminutive forms (e.g., Bobbie). It may be said at once that male first names in American English very seldom occur in full form (Robert, James, or Gerald) but are almost always either abbreviated (Bob, Jim) or diminutized (Jerry) or both (Bobbie, Jimmy). Female first names are more often left unaltered. Titles for the purpose of this classification include, in addition to Mr., Mrs., and Miss, such occupational titles as Dr., Senator, Major, and the like.

Two Reciprocal Patterns

The vast majority of all dyads in the plays exchange FN (Mutual FN). Indeed, where the actual name is not known there occur sometimes what may be called generic first names; these include the Mack, Jack, and Buddy of taxi drivers. Mutual TLN is most commonly found between newly introduced adults. The

[2] Since address involves two persons and, very often, a choice between two linguistic forms, we believe it should be possible to treat the complete set of address dyads for any group of persons as a matrix of paired comparisons from which the dimensions structuring the group could be rigorously derived. We are at present trying to develop such a method.

distinction between the two patterns is primarily one of degree of acquaintance with the degree required for the Mutual FN being less for younger people than for older people and less where the members of the dyad are of the same sex than where they are of different sex.

It seems likely that the two reciprocal patterns are on a dimension that ranges from acquaintance to intimacy. However, in modern American English the distance between the two points is small with the Mutual FN usually representing only a very small increment of intimacy over the Mutual TLN; as small sometimes as 5 minutes of conversation. Because the segment of the line that lies between the two patterns is usually so very short, it is not easy to make out its exact character. However, in English of the past[3] and in cognate languages today the Mutual FN is farther displaced from the Mutual TLN, and from these cases and other materials later to be presented we can hazard a characterization of intimacy. Intimacy is the horizontal line between members of a dyad. The principal factors predisposing to intimacy seem to be shared values (which may derive from kinship, from identity of occupation, sex, nationality, etc., or from some common fate) and frequent contact. Among the behavioral manifestations of intimacy, a relatively complete and honest self-disclosure is important.

Nonreciprocal Pattern

In this case one member of the dyad says FN and the other TLN. There are two kinds of relation that can generate this pattern. The first is a difference of age: children say TLN to adults and receive FN; among adults an elder by approximately 15-or-more years receives TLN and gives FN to his junior. The second is a difference of occupational status: this may be a relation of direct and enduring subordination (e.g., master-servant, employer-employee, officer-enlisted man); it may be a relation of direct but temporary subordination, involving someone in a service occupation (e.g., waiter, bootblack) and a customer; it may be an enduring difference of occupational status that does not involve direct subordination (e.g., United States senators have higher status than firemen). If the intimacy dimension that governs reciprocal address is the horizontal of social relationship, then the status dimension that underlies the nonreciprocal pattern may be called the vertical of social relationship.[4]

Age and occupational status are correlated and most instances of nonreciprocal address involve congruent differences on the two dimensions. There is, however, proof that a difference on either dimension alone is able to generate the nonreciprocal pattern. The proof is the existence of nonreciprocal dyads matched on one dimension but not on the other.

Because there are two criteria for the assignment of status and, in addition, the correlation between the two criteria is not perfect, we must ask what happens to address in dyads where the elder has the humbler occupation. There are in the plays numerous instances in which the criteria oppose one another: an adolescent girl and her family's middle-aged cook, a young navy ensign and a middle-aged enlisted man, a young executive and an elderly janitor. In all such cases address is in accordance with occupational status and so it would appear that there is a normative rule of priority for the two criteria. It is to be expected in a society whose values are more strongly linked to achieved personal attributes than to ascribed attributes (Parsons, 1951) that occupation would prevail over age in the determination of deference.

The three sets of data in addition to the plays confirm the generalizations made above. Sloan Fellows call almost all of their business acquaintances by FN and expect this address to be reciprocated. In the few cases where one of these men would say TLN and expect to receive FN the other member of the dyad was invariably an organizational superior and was also the elder. In the Boston drafting firm, also, most address is on the pattern of Mutual

[3] In six American plays (Quinn, 1917) written between 1830 and 1911 the reciprocal FN between adults clearly implies a much longer and closer acquaintance than it does in contemporary usage.

[4] Kinship terms of address in American English (Schneider & Homans, 1955) also show a nonreciprocality of status. Members of ascending generations are commonly addressed with kinship titles (mother, father, grandmother, grandfather, uncle, aunt) but respond by calling their children, grandchildren, nephews, and nieces by FN.

FN. The few cases of Mutual TLN involved persons who were scarcely acquainted and who were well matched by age and position. There were 40 different instances of nonreciprocal address. In 36 of these the recipient of TLN was the organizational superior and elder; in 28 cases he was a member of one of the four top executive ranks whereas the recipient of FN was a member of one of the eight unionized ranks. In three cases the organizational superior was younger than his subordinate and it was the superior, not the elder, who received TLN. The single remaining case involved a pair matched by rank but not by age and in this case the elder received TLN. Midwest children participate in nonreciprocal address dyads with parents and teachers and in Mutual FN dyads with other children.

VARIANT FORMS OF ADDRESS

In this section are offered tentative characterizations of the several common forms other than FN and TLN. At a later point in the paper we will return to these variants and suggest the nature of their relationship with the major patterns.

Title Without Name

Commonly used titles (T) include sir, madam, ma'am, and Miss. In general these forms are used like TLN; either reciprocally between new acquaintances or nonreciprocally by a person of lower status to a person of higher status. The address form T is probably a degree less intimate and a degree more deferential than TLN. It may, for instance, be used reciprocally where acquaintance is so slight that the last name is not known. In nonreciprocal military usage between ranks the TLN may be used to immediate superiors but the T to remote colonels, generals, commanders, and admirals even though the names of these superiors are well known.

Beyond this general characterization as a minimally intimate and maximally deferential form particular varieties of T have their specialized uses. The form ma'am is most commonly heard from young men to mature women. Schoolchildren in Yoredale, England preface almost all address to a teacher by saying: "Please, Miss . . . " In parallel circumstances Midwest children use TLN.

Last Name Alone

An occasional person is regularly addressed by LN. This seems usually to occur where the FN is polysyllabic and has no familiar abbreviation whereas the LN is either a monosyllable or readily transformed into a monosyllable. In these circumstances LN is simply a substitute for FN and patterns in identical fashion.

Where the LN is not the usual form for an addressee it represents a degree of intimacy greater than TLN but less than FN. In military usage enlisted men receive the LN from officers when they are little acquainted; increased familiarity leads to the FN downward though not upward. Elderly and very distinguished professors sometimes begin letters to junior colleagues whom they know fairly well: "Dear Jones . . . " The form is not reciprocated in this case.

Reciprocal LN is common between enlisted men until they become acquainted. The enduring reciprocal LN seems always to go with a mutual antagonism that blocks progression to intimacy. In *The Caine Mutiny Court Martial*, for instance, Greenwald and Keefer exchange LN though the degree of acquaintance and their ages would normatively produce Mutual FN. In the climax of this play Greenwald dashes a drink in Keefer's face. In *The Philadelphia Story* Kittredge and Haven exchange LN; the former is the husband-to-be of Tracy Lord and the latter her former-but-still-interested husband.

Multiple Names

A speaker may use more than one form of the proper name for the same addressee, sometimes saying TLN, sometimes FN or LN or a nickname, sometimes creating phonetic variants of either FN or the nickname. We are not here interested in the business of temporal progression through the possible forms where earlier terms are dropped as new terms are taken up. The case of multiple names (MN) is the case in which two or more versions of the proper name are used in free variation with one another.

The instances of MN in the plays suggested that this form represented a greater degree of intimacy than the FN, but degrees of intimacy are not easily judged and so we decided to put the hypothesis to a more direct test. The

test involved individual interviews with 32 MIT male undergraduates. Each subject was asked to think of four men of about his own age all of whom he had met for the first time approximately one year ago. In addition, it was to be the case that the subject had had about equal opportunity to get to know all of these men and yet now was to find himself on close, friendly terms with some of them and on more distant terms with others. The subject was asked to write down the full names of the four men and then to record the name by which he usually addressed each one. When this had been done, the subject was asked, for each acquaintance, whether he ever addressed the man in any other way and all the names currently being used were noted down.

In an earlier paragraph it was suggested that one of the meanings of intimacy is a relatively complete and honest disclosure of the personality. Jourard and Lasakow (1958) have devised a Self-disclosure questionnaire which requires the subject to indicate whether or not he has discussed with a designated other person each of 60 topics classifiable under the headings: Attitudes and opinions, Tastes and interests, Work (or studies), Money, Personality, and Body. Our 32 subjects filled out this questionnaire for each acquaintance. For each subject, then, we ranked the four acquaintances in order of decreasing Self-disclosure and we take this to be an order of decreasing intimacy. It generally corresponds with the subjects' ordering of the four into friends and acquaintances.

Across all subjects we combined the 32 closest friends, those eliciting highest Self-disclosure scores (Self-disclosure 1) and also the 32 acquaintances with Self-disclosure rank orders 2, 3, and 4. For each rank group we determined the number of cases addressed by MN (2 or more proper names) and the number addressed by FN alone. The results appear in Table 1. The cases of MN decline as intimacy declines. Using the sign test and a one-tailed hypothesis it was determined that the Self-disclosure 1 group contained more cases of MN than did the next most intimate group (Self-disclosure 2) with $p = .020$. The difference between the most intimate and the least intimate (Self-disclosure 4) group shows more cases of MN in the former group with $p < .004$.

TABLE 1

FREQUENCIES OF TWO FORMS OF ADDRESS TO ACQUAINTANCES AT DIFFERENT LEVELS OF SELF-DISCLOSURE

	Self-disclosure 1	Self-disclosure 2	Self-disclosure 3	Self-disclosure 4
Multiple Names	18	10	7	3
First Name	14	22	25	29

Note.—Self-disclosure 1 vs. Self-disclosure 2 by sign test, $p = .020$; Self-disclosure 1 vs. Self-disclosure 4 by sign test, $p < .004$.

One informant addressed his closest friend whose name is Robert Williams as Williams or Robert or Bob or Willie and his next closest friend whose name is James Scoggin as Scoggin, James, Jim, or Scoggs. Many informants reported that they sometimes playfully addressed a good friend by TLN. Others used playful, and usually pejorative, phonetic variations: Magoo for Magee, Katool for Katell, Lice for Leis.

The tendency to proliferate proper names in intimacy is interesting because it accords with a familiar semantic-psychological principle. For language communities the degree of lexical differentiation of a referent field increases with the importance of that field to the community. To cite a fresh example of this kind of thing, Conklin (1957) reports that the Hanunóo of the Philippine Islands have names for 92 varieties of rice which is their principal food. In naming ferns and orchids, with which they are little concerned, the Hanunóo combine numerous botanical species under one term whereas the rice they differentiate so finely is for the botanist a single species. Within a language community Brown (1958) has pointed out that a speaker more concerned with a given referent field will make finer lexical distinctions than a speaker less concerned with that field (botanists have more names for plants than do psychologists). In the referent field composed of other persons we have seen that where contact and concern are minimal and distance greatest, titles alone are likely to be used in address. To call someone Miss or sir is to address the person on a categorical level which does not establish the addressee's individual identity. The proper name constitutes the individual as a unique organism. Beyond the single proper name, however, where interest is still greater the

individual is fragmented into a variety of names. Perhaps this differentiation beyond individuality expresses various manifestations or ways of regarding someone who is close (Brown, 1959).

A General System of Address

How are these various forms of address related to one another? Consider first only the three major patterns. The Mutual TLN goes with distance or formality and the Mutual FN with a slightly greater degree of intimacy. In nonreciprocal address the TLN is used to the person of higher status and the FN to the person of lower status. One form expresses both distance and deference; the other form expresses both intimacy and condescension. Within the limits of two dyadic address forms there is a formally or logically possible alternative pattern. The form used mutually between intimates could be used upward to superiors and the form used between distant acquaintances could be directed downward to subordinates. Because there is an alternative the pattern actually found is not a formal necessity but rather an empirical fact; a fact, we shall see, of great generality.

Several years ago we began our general studies of the language of social relationship with a selection point that does not occur in modern English but which does occur in the other Indo-European languages. In French, for example, a speaker must choose between two second person singular pronouns; his addressee may be addressed as *tu* or as *vous*. In German the comparable forms are *du* and *Sie;* in Italian *tu* and *Lei*. In English of the past, from about the 13th century until the 18th, there was a cognate option of address; the choice between *thou* and *ye*. We have studied the semantic rules governing these pronouns in 20 languages of Europe and India, comparing one language with another and the usage of earlier centuries with later (Brown & Gilman, 1960). For our present purposes the important point is that these pronouns in all the languages studied follow the same abstract pattern as the FN and TLN.

In discussing the pronouns of address let us use T as a generic designator for pronouns of the type of *tu* and *du* and V as a designator for pronouns of the type of *vous* and *Sie*. Mutual V is the form of address for adult new acquaintances; it begins where TLN begins—at the temporal point of origin of the dyad. Mutual T like Mutual FN is an expression of increased intimacy but it is, for most Europeans, much farther along the line than is the Mutual FN of Americans. From medieval times into the present century nonreciprocal T and V was the pattern for those unequal in status with the superior receiving the V and the subordinate the T. In recent times the nonreciprocal use of the pronouns has much declined because of a conscious egalitarianism. The important point is, however, that when the nonreciprocal pattern has been used anywhere from southern India to Scandinavia the downward directed form has been the intimate T and the upward directed the distant V.

It may be that the abstract linkage in personal address of intimacy and condescension, distance and deference is a linguistic universal, but we certainly do not know that as yet. We do know that the linkage occurs also in some non-Indo-European languages (e.g., Japanese second person pronouns). Indeed, in those few languages for which we have found adequate descriptions of the semantics of address, no violations of the abstract pattern have yet appeared.

It seems also that the pattern applies to more than names, titles, and pronouns. The Sloan Fellows' greetings to their business associates almost invariably used the FN. However, the greetings themselves were quite varied; including Hi, Morning, Good Morning, Hello, Howdy, etc. Only Hi and Good Morning occurred with sufficient frequency to make possible the discovery of a pattern. The Sloans reported on greetings to four classes of associate: Equal and Intimate, Equal and Distant, Superior, and Subordinate. In Table 2 appear the frequencies of Hi and Good Morning for these various categories. Hi is more common to intimates and to subordinates while Good Morning is for distant acquaintances and superiors. Using the McNemar test for the significance of changes in related samples with the Yates correction for continuity, the difference between intimates and acquaintances is significant with $p < .0025$ and the difference for subordinates and

TABLE 2
TWO FORMS OF GREETING FOR FOUR CLASSES OF ASSOCIATE

	Equal and Intimate	
	Good Morning	Hi
Equal and Distant		
Hi	0	4
Good Morning	4	10
	Subordinate	
	Good Morning	Hi
Superior		
Hi	1	3
Good Morning	13	8

Note.—With McNemar test for the significance of changes in related samples (employing Yates correction) χ^2 for Intimate vs. Distant = 8.10, $p < .0025$ and χ^2 for Subordinate vs. Superior = 4.00, $p < .025$.

superiors with $p < .025$. In both cases a one-tailed test was used since the direction of the differences was predicted by the abstract pattern of address. The records of actual usage in a Boston firm accord with this finding. In one revealing instance a workman was greeted "Hi" and promptly answered "Hi," but as he turned and recognized the boss, he added "Good morning."

Why should the abstract pattern described govern address between two persons? A curious fact about contemporary use of T and V provides a clue. While the nonreciprocal pattern for pronouns has generally been abandoned in Europe, inequality of status continues to affect one aspect of usage. Dyads begin at the Mutual V and, with time, may advance to the intimacy of Mutual T. For many the shift from V to T is an important rite of passage. The Germans even have a little informal ceremony they call the *Bruderschaft*. One waits for a congenial mood, a mellow occasion, perhaps with a glass of wine, and says: "Why don't we say *du* to one another?" The new usage is, of course, to be reciprocal. However, there is one necessarily nonreciprocal aspect of the occasion—someone must make the suggestion. When there is a clear difference of status between the two the right to initiate the change unequivocally belongs to the superior—to the elder, the richer, the more distinguished of the two. The gate to linguistic intimacy is kept by the person of higher status.

The norms of English address also make a pattern in time. A dyad must, with time, either increase its total amount of contact or else dissolve. Since the Mutual TLN represents less contact than the Mutual FN if Mutual TLN is to occur in a given dyad it must occur at an earlier time than the Mutual FN. The place of the nonreciprocal pattern in time is between the other two and it may be understood as a step from Mutual TLN in the direction of Mutual FN; a step which, like the suggestion of the *Bruderschaft* is taken first by the superior. Many dyads will linger for a very long time—possibly the life of the dyad—in the nonreciprocal pattern. In this circumstance the pattern gives enduring expression to an inequality of status.

Consider a familiar sort of example. A prospective graduate student arrives at a university to meet some of the faculty of the psychology department and is interviewed by the chairman. Probably the two will initially exchange TLN. In the course of the day or, if not, shortly after the student enrolls, the chairman will begin to call him by FN. He extends the hand of friendship, but the student knows that it behooves him not to grasp it too quickly. The student will continue with the TLN for several years (4 is probably the mode) and in this period the nonreciprocality of speech will express the inequality of status. If the chairman is neither very elderly nor very august the student will eventually feel able to reciprocate the FN and the dyad will have advanced to Mutual FN. The three patterns may be described as a progression in time (see upper portion of Figure 1) if we add several important qualifications.

Not every dyad passes through all three steps. There are some that begin at each of the three points: adults of equal status with Mutual TLN, master and servant with nonreciprocality, young people with Mutual FN. In addition, not every dyad that continues to exist will necessarily advance to Mutual FN. In North Carolina until 1860 the Negro slave said TLN to his master and was told FN (Eliason, 1956), and there was no change with time. There is a final qualification concerning the progression of address in time. Even when

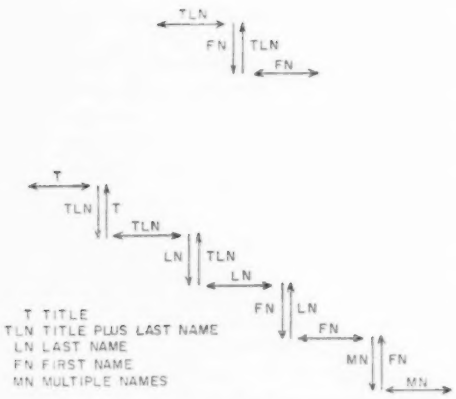

FIG. 1. Graphic models of the progression of address in time (from left to right). (The upper portion of the figure represents the major progression; the lower portion represents the full progression.)

relationships do develop in intimacy they will not necessarily pass through the intermediate nonreciprocal stage. When the members of a dyad are not of clearly unequal status, they will advance at the same time to Mutual FN.

The general statement that can be made is: if an address pattern changes in time it will change in the direction of Mutual TLN → nonreciprocal TLN and FN → Mutual FN though a step may be skipped. Even this statement cannot stand without a little more explanation. There are special circumstances in which the direction of movement of address can be reversed. If a person of lower status seems to move too fast to the reciprocation of the FN a superior may step back from his use of FN to TLN. The day after a convivial office party a breezy young clerk calls out to the president: "Morning, Jack!" and in icy tones the president replies: "Good morning, *Mr.* Jones." The person of lower status must never use a more familiar form of address than the person of higher status and the backstep by the superior puts the subordinate in just that position from which he will usually withdraw to the propriety of TLN.

The variant forms of address discussed at an earlier point in this paper seem to function as additional optional steps in the progression of address.[5] Title alone (T) probably is the formal extreme since we find in our data that dyads which begin as Mutual T often change with time to nonreciprocal TLN downwards and T upwards. The last name alone (LN) must be intercalated between TLN and FN since LN downwards is found in combination with TLN upwards and LN upwards is found in combination with FN downwards. The use of multiple names (MN), we have seen, represents the intimate extreme. Making these additions to the major three-step progression we arrive at the full progression of Figure 1 (lower portion) in which each new step towards intimacy is initiated by the superior who is, therefore, not just the gatekeeper to Mutual FN but the pacesetter for all linguistic advances in intimacy.

The qualifications that apply to the major progression apply also to the full progression. Dyads may begin at any point, need not move at all, and if they move may skip steps. There is an additional qualification that results from the multiplication of forms. Any less formal term may be used downwards in a nonreciprocal pattern with any more formal term being used upwards. Thus we find not only the TLN with the LN but also the T with LN or even with FN as when a captain calls an enlisted man *Jones* or *John* but is called *captain*.

Is there any way to test the accuracy of the full progression pictured in Figure 1? We have found no dyads that pass through the entire progression but only dyads moving in one or another limited region. By the time the third and final set of plays was ready for analysis, the full progression had been constructed. If the progression, with all its qualifications, is accurately descriptive of American English practice, one should find certain kinds of address combinations and not others. Any address form can be used mutually (5 possible combinations). When there is a clear difference of status we may have any nonreciprocal pattern that combines a less formal term downwards with a more formal term upwards (10 possible combinations). With a clear status difference we may not have any nonreciprocal pattern that combines a more formal term

[5] This paper does not discuss all American English forms of address but only those that have clear positions in the intimacy and status pattern. There is, for instance, the use of the complete name (*John Jones* or even *John Montgomery Jones*) which is used as an intensifier and is particularly favored by mothers "manding" disobedient children.

downwards with a less formal term upwards (10 impossible combinations). In the last set of plays instances of all the possible combinations occur and no instances of the impossible combinations. The static predictions of the model are validated.

The construction also generates a set of dynamic predictions. When address changes (if there is not either a reproving or a jocular intent) movement must be in the direction from left to right in the drawing of Figure 1 though it need not be to the immediately adjacent position. Changes of address in the plays are surprisingly infrequent and seldom involve more than two steps. Only seven different kinds of change occurred but all of these are included among those defined as possible by the model. One "impossible" change occurred in the play *Born Yesterday*. Billie Dawn, the junk dealer's mistress, in speaking with the wife of a United States Senator who is many years her senior initiates movement from Mutual TLN to a more intimate form. This violation of the norms is greeted with general shock and is a device that helps to establish Billie Dawn's ingenuous vulgarity. Insofar as the limited occurrences of change in address permit, then, the model is confirmed in its dynamic aspect.

STATUS AND INTIMACY IN SOCIAL RELATIONS

Two persons of unequal status may be conceived as two points on a generalized value scale of the sort used by Osgood and Tannenbaum (1955) for the congruity model which they have used to predict attitude change. The person of superior status has, of course, the greater value on such a scale. Movements towards intimacy of address in terms of the congruity model are acts of association. Such acts, the model predicts, will cause the objects of unequal value to move towards one another; in the case of address between persons this means that the value of the inferior is enhanced by intimate association with a superior while the value of the superior is diminished. The prediction is good intuitive sense. But now the interesting thing is that this model seems to call for the abstract pattern of address that does not occur. Since the person of lower status has a motive for initiating intimacy and the superior has none the intimate form ought, in nonreciprocal address, to be used upwards and the distant form to be used downwards. This is not the pattern which we have found in all of our materials but, is rather, the formal alternative to that pattern and we have nowhere found this alternative to be the operative norm.

The abstract design of address is not a direct expression of the realities of status and intimacy but is rather a denial of the realities. The pattern might have been designed to minimize the pain of invidious status distinctions. The person of lesser value may be presumed to be ever ready for association with a person of higher value but the person of higher value must be presumed to be less ready. If the person of lesser value were to initiate associative acts, he would run the risk of rebuff; if the person of higher value initiates such acts there is no such risk. The superior, then, must be the pacesetter in progression to intimacy. If there is to be no progression but rather an enduring nonreciprocal expression of inequality, this expression is not so disagreeable as it might be since the superior offers intimate address and it is the inferior who demurs. The abstract pattern minimizes the pain that could be involved when persons are to be related on two dimensions: the horizontal of intimacy and the vertical of status.

The person of higher status is, we believe, the pacesetter not in linguistic address alone but in all acts that increase intimacy. The Sloan Fellows who served as informants for us responded to two questions that concern nonlinguistic moves towards intimacy. Each

TABLE 3

COMPARATIVE WILLINGNESS TO INITIATE ACTS OF INTIMACY WITH FOUR CLASSES OF ASSOCIATE

	More Willing with Equal and Intimate Than with Distant Associate	More Willing with Equal and Distant Than with Intimate Associate	More Willing with Subordinate Than with Superior	More Willing with Superior Than with Subordinate
Borrow Comb	7	0	9	2
Hand on Shoulder	11	3	24	1

Note.—By sign test Intimate vs. Distant on Borrow Comb, $p = .008$; on Hand on Shoulder, $p = .029$. Subordinate vs. Superior on Borrow Comb, $p = .033$; on Hand on Shoulder, $p < .002$.

question was answered with reference to each of the four persons listed by the informant. One of these persons, you may remember, was a Superior, one a Subordinate, one Equal and Intimate, one Equal and Distant. The informant indicated on a scale from 1–5 how willing he would be to behave in the designated manner with reference to each person. The Number 1 represents maximal willingness ("Definitely would") and 5 minimal willingness ("Definitely would not"). The questions are:

1. In ordinary circumstances would you be willing to ask "X" for the loan of his comb?
2. Suppose "X" were feeling very unhappy about something or other. Would you put your hand on his shoulder in a reassuring way?

The results appear in Table 3 and show a familiar pattern. Informants were more willing to borrow a comb from, and to put a hand on the shoulder of, an associate who was Equal and Intimate than one Equal and Distant and also more willing to initiate these acts of intimacy with Subordinates than with Superiors. Using a sign test and a one-tailed hypothesis the comb question differentiates intimate from remote associates with a $p = .008$ and Subordinates from Superiors with a $p = .033$; the hand-on-shoulder question differentiates intimate from remote associates with a $p = .029$ and Subordinates from Superiors with a $p < .002$. These nonspeech acts follow the pattern of TLN and FN; of T and V; and of Hi and Good Morning. Perhaps all kinds of associative behavior can be placed on a rough scale of intimacy and in the progression over this scale of a dyad the superior may always be in advance.

With a good sized inequality of status the use of FN by the higher does not, of course, justify immediate reciprocation from the lower. The lower must wait for the initiation of additional acts of intimacy before taking a step on his own. It is very likely that the normative lag in intimacy increases with the degree of status inequality. However, the norms are not always perfectly clear; graduate students will sometimes be uncertain whether the time has come to say FN to a professor, employees will wonder whether they know their bosses well enough to use the familiar form. When someone is in this region of uncertainty, we find that he avoids the use of any sort of personal name and makes do with the uncommitted omnibus *you*.

SUMMARY

The semantic rules governing address in American English are worked out from a varied collection of data that includes usage in American plays, actual usage in a Boston business firm, reported usage of business executives from various cities in the United States, and usage recorded in a midwestern American town. The most common address forms are the first name (FN) and the title plus last name (TLN). These function in three sorts of dyadic pattern: the Mutual TLN, the Mutual FN, and the nonreciprocal use of TLN and FN. The semantic distinction between the two mutual patterns is on the intimacy dimension with Mutual FN being the more intimate of the two patterns. In the nonreciprocal pattern a distinction is made in terms of status with the higher saying FN and the lower TLN. The practice of using the intimate form to a subordinate and the distant form to a superior also governs the use of pronouns of address in many languages as well as the use of certain conventional greetings. It is suggested that this very general pattern prevails because in the progression towards intimacy of unequals the superior is always the pacesetter initiating new moves in that direction. The superior is the pacesetter because the willingness of the person of lower status to enter into association can be taken for granted and there is little risk that a superior will be rebuffed whereas the risk would be great if the inferior were to initiate acts of association. Such variant forms of address as the title alone, the last name alone, and the use of multiple names are fitted into a model that purports to describe the temporal progression of address from acquaintance to friendship. Each new step towards friendship is, in this model, initiated by the person of higher status.

REFERENCES

BARKER, R. G., & WRIGHT, H. F. *Midwest and its children*. Evanston, Ill.: Row, Peterson, 1954.

BROWN, R. *Words and things*. Glencoe, Ill.: Free Press, 1958.

BROWN, R. Humbert's idiography. *Contemp. Psychol.*, 1959, **4**, 172–174.

BROWN, R., & GILMAN, A. The pronouns of power and solidarity. In T. Sebeok (Ed)., *Style in language*. New York: Wiley, 1960.

CONKLIN, H. C. *Hanunóo agriculture*. Rome: Food and Agriculture Organization of the United Nations, 1957.

ELIASON, N. E. *Tarheel talk: An historical study of the English language in North Carolina to 1860*. Chapel Hill: Univer. North Carolina Press, 1956.

GASSNER, J. (Ed.) *Best plays of the modern American theatre: Second series*. New York: Crown, 1947.

GASSNER, J. (Ed.) *Best American plays: Third series*. New York: Crown, 1952.

GASSNER, J. (Ed.) *Best American plays: Fourth series*. New York: Crown, 1958.

JOURARD, S. M., & LASAKOW, P. Some factors in self-disclosure. *J. abnorm. soc. Psychol.*, 1958, **56**, 91–98.

OSGOOD, C. E., & TANNENBAUM, P. H. The principle of congruity in attitude change. *Psychol. Rev.*, 1955, **62**, 42–55.

PARSONS, T. *The social system*. Glencoe, Ill.: Free Press, 1951.

QUINN, A. H. (Ed.) *Representative American plays*. New York: Century, 1917.

SCHNEIDER, D. M., & HOMANS, G. C. Kinship terminology and the American kinship system. *Amer. Anthrop.*, 1955, **57**, 1194–1208.

(Received March 21, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 386-390

WHAT DO ATTITUDE SCALES MEASURE:

THE PROBLEM OF SOCIAL DESIRABILITY

JAMES BENTLEY TAYLOR[1]
University of Washington

MUCH recent research on personality measurement indicates that test-taking behavior is highly influenced by group norms of social desirability. A number of studies have reported that the more socially desirable a statement is, the more apt it is to be endorsed. There is also evidence to show that people differ in their willingness to endorse socially undesirable statements. This difference accounts for much of the variance of scores on personality tests, even when the tests supposedly measure quite different aspects of personality (Edwards, 1957).

Something similar may take place with attitude scale items. A person who tends to give a socially desirable impression of himself may also tend to endorse only those items on an attitude scale which he perceives as socially desirable. Such a response set might thus influence a person's scores on a number of attitude scales, producing "spurious" interscale correlations. For instance, a person who scores high on an "authoritarianism" scale may also score high on a scale measuring "anti-Semitism" because of a general willingness to endorse socially undesirable attitude items.

This supposition implies the following four hypotheses:

1. Social desirability norms exist for attitude-scale items.

2. People differ in the extent to which they show a general response tendency to endorse socially undesirable attitude items.

3. A measure of such a general tendency should be highly correlated with scores on a variety of attitude scales.

4. People who tend to endorse a large number of socially undesirable *personality* statements also tend to endorse a large number of socially undesirable *attitude* statements.

[1] Thanks are due to Walter Gerson, William Larson, and George Rothbart for their aid in providing subjects, and to Judith Nelson for computational assistance.

The present paper reports an investigation of these four hypotheses.

But even if these hypotheses are confirmed, acquiescence response set, as originally identified by Lentz (1938), must be considered as an alternate explanation of the findings. A final hypothesis is therefore investigated:

5. Assuming the confirmation of Hypotheses 1–4, the results cannot be explained in terms of acquiescence response set.

Three studies were conducted, the first testing Hypothesis 1, the second Hypotheses 2 and 3, and the third testing Hypotheses 4 and 5. These studies are reported separately below.

STUDY I: SOCIAL DESIRABILITY AND ITEM ENDORSEMENT

Method

A pool of 127 attitude statements, from six attitude scales, was collected for the first study. Included were 26 items from the California F Scale (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950), a 5-item Anomie scale (Srole, 1956), a 20-item scale of Ethnocentrism (Adorno et al., 1950), a scale of Radicalism-Conservatism with 15 items (Eysenck, 1955), and 50 items drawn from the Parent Attitude Research Instrument—hereafter referred to as the PARI (Schaefer & Bell, 1958). Ten scales from the PARI were included but were not analyzed as separate scales, since factor analysis of the PARI has shown factor loadings on two dimensions: "Approval of maternal control of child," and "Approval of expression of hostility" (Schaefer, 1957). The PARI items were scored for these two factor scales. In addition, nine "filler" items were included in the pool of attitude statements. All 127 items were rated by 53 undergraduates at the University of Washington on a nine-point scale of social desirability, using instructions adapted from those given by Edwards (1957). The same 127 items were given to another group of students ($n = 58$), with the request to agree or disagree with each item.[2]

[2] An 8-page table giving the instructions for the social desirability ratings, and the median social desirability rating for the 127 attitude scale items, has been deposited with the American Documentation Institute. Order Document No. 6743 from ADI Auxil-

The median social desirability rating, the probability of item endorsement, and the interquartile range was computed for each item. The median social desirability ratings were then correlated with the item endorsement values. In addition, each of the six scales in the item pool was analyzed separately to find the average median social desirability of its items, the mean interquartile range of the items, and the standard deviation of the median social desirability ratings.

Results

The average social desirability ratings for scale items, the average interquartile ranges of the items, and the standard deviations of the medians are reported in Table 1. The "average rating" shows the extent to which the scale items are generally weighted with social desirability, the average Q values indicate the degree of consensus on social desirability for items in the scale, and the standard deviation of medians gives an index of the variability between scale items on the social desirability continuum. It will be noted that all of the scales are loaded with socially undesirable items, with the exception of the PARI scale for "Approval of Hostility." Of the total 127 scale items, only 13 had a median rating on the socially desirable half of the scale (a score of 5 or above). When the median social desirability ratings were correlated with the probability of item endorsement, a correlation of .80 was obtained.

A number of statements in the item pool clearly implied that socially desirable results would follow from certain acts; agreement with such a statement in effect assigned high social desirability to it. The incorporation of such statements might be expected to result in a spuriously high correlation between item endorsement and social desirability values. Therefore, the probability of item endorsement was correlated with the median social desirability values of the items with these "social action" statements removed. The correlation was .74.

Of the six scales, five were scored only in the socially undesirable direction. That is, if a person answered in a socially undesirable direction, he was scored as being higher in whatever the scale measured (Ethnocentrism, Anomie, etc.). The sole exception was the Eysenck Radicalism Scale, in which 7 of the 15 items were scored so that a socially desirable response added to the scale score. It might be expected that this "balancing out" of social desirability would make a score on the Radicalism scale less influenced by the social desirability variable.

TABLE 1
MEDIAN SOCIAL DESIRABILITY RATINGS, STANDARD DEVIATIONS, AND INTERQUARTILE RANGES OF ITEMS FOR SIX ATTITUDE SCALES

Scale	Average Median Rating	σ of Median Ratings	Mean Item Q
Authoritarian	3.34	1.42	3.4
Anomie	2.48	2.32	2.8
Ethnocentrism	2.28	1.18	2.9
Radicalism	2.94	1.41	2.9
PARI Control	3.23	1.12	3.3
PARI Hostility	4.59	.99	3.4

From these findings it is apparent that definite social desirability norms exist for these attitude-scale items; that, with one exception, the scales which make up the pool are loaded with socially undesirable items; and that the high correlation between probability of item endorsement and social desirability of items which has been demonstrated for personality statements also characterizes attitude-scale items.

STUDY II: GENERALITY OF SOCIAL DESIRABILITY RESPONSE SET

Method

In order to test the generality and significance of the social desirability set on attitude-scale scores, a Social Desirability scale was devised. From the pool of items collected in the first study, 59 with a median social desirability value of 3.0 or less were selected as being least socially desirable. The 59 statements were divided at random into two groups, of 29 and 30 items. It will be recalled that 58 students had responded to the items in the statement pool in terms of agreement or disagreement. Each of these students was assigned two scores: the number of items endorsed in the first group of statements, and the number of items endorsed in the second, and a correlation between scores computed. If the students were indeed differentially influenced by a general set to give socially desirable responses, then we would expect a significant and high correlation.

If a general set of social desirability influences attitude-scale scores, the Attitude Social Desirability score should correlate substantially with scores on the

iary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

other scales. It was not, of course, possible to use the entire Attitude Social Desirability scale in each correlation, since it included items from each of the six attitude scales. In comparing scores on Social Desirability with scores on Authoritarianism, and again when comparing Social Desirability with Ethnocentrism, *both* the Authoritarianism and the Ethnocentrism items were dropped from the Social Desirability scale. Similarly, in correlating each of the two PARI scales with the Social Desirability scale, all PARI items were dropped from the latter. In the case of correlations with Radicalism and with Anomie, the overlapping items were omitted from the Attitude Social Desirability scale. All correlations were Pearsonian *r*'s, with the exception of the correlation between the Attitude Social Desirability and the Anomie scale, which consisted of only five items with a very limited response range. In this case the subjects were divided into those who endorsed at least one item on the Anomie scale, and those who endorsed none, and biserial coefficient of correlation was computed for these two groups.

Results

When the split-half test of reliability was made between the two groups of socially undesirable items, a correlation of .74 was obtained which, by the Spearman Brown Prophecy formula, corresponds to a reliability coefficient of .85.

The correlations between the partial Attitude Social Desirability scale and the other scales are given in Table 2. As can be seen, there is a substantial correlation between Attitude Social Desirability and the scale scores for all except one scale. The single exception, Eysenck's Radicalism Scale, is also the only scale in which the items are scored so as to balance out social desirability.

These results indicate that people differ in the extent to which they endorse socially undesirable attitude items, that this difference in willingness is a general response tendency over several types of attitudes or items, and that the tendency to endorse socially undesirable attitude statements is strongly correlated with scores on a number of attitude scales that are diverse in content.

STUDY III: ATTITUDINAL AND PERSONAL SOCIAL DESIRABILITY

Method

In this third study two additional measures of social desirability set were developed, besides the scale of Attitude Social Desirability just described: the "Personal Social Desirability" scale, and the "Reversed Social Desirability" scale. As previously reported (Taylor, 1959), university students had rated

TABLE 2
CORRELATIONS BETWEEN A PARTIAL ATTITUDE SOCIAL DESIRABILITY SCALE AND SIX ATTITUDE SCALES

Scale	Correlation
Authoritarian	.65
Anomie	.57[a]
Ethnocentrism	.61
Radicalism	−.21
PARI Control	.64
PARI Hostility	.42

[a] Biserial coefficient; other values are product-moment.

TABLE 3
INTERCORRELATIONS BETWEEN THREE SOCIAL DESIRABILITY SCALES

Scale	Attitude SD	Personal SD	Reversed SD
Attitude SD	.85[a]	.58	.32
Personal SD		.83[a]	.63
Reversed SD			.50[a]

[a] Split-half estimates of total scale reliability.

205 MMPI items on a nine-point scale of social desirability. On the basis of these ratings, 59 MMPI statements that had been rated as very undesirable (i.e., given a median rating of three or less) were selected to make up a scale of Personal Social Desirability. A high score on this scale was held to measure the degree to which a person was apt to endorse socially undesirable personality statements. In addition, 38 of the MMPI items that had been given median ratings of seven or above on the socially desirable end of the continuum were grouped into another scale, entitled Reversed Social Desirability. A high score on this scale measured the degree to which a person was willing to *deny* favorable statements about himself.

These three scales—Attitude Social Desirability, Personal Social Desirability, and Reversed Social Desirability—were then given to 50 undergraduate students at the University of Washington. The reliability for each scale was estimated, using the split-half technique and estimating the total reliability with the Spearman Brown Prophecy formula. The three scales were also intercorrelated.

Results

Table 3 lists the obtained intercorrelations and the estimated reliabilities of the three scales. As can be seen, a substantial relationship was found between willingness to endorse socially undesirable *attitude* statements and willingness to endorse socially undesirable *personality* statements. Corrected for attenuation, the correlation is .70.

There was a substantial relationship between willingness to *endorse* socially undesirable personality items and willingness to *deny* socially desirable items, contrary to

what would be expected if the social desirability scales were simply measuring an acquiescence response set. This finding is similar to that reported by Edwards (1957) in other studies of the social desirability set.

These findings indicate that people who tend to endorse a large number of socially undesirable attitude statements also tend to endorse a large number of undesirable personality statements; and that the findings reported in this paper can best be explained in terms of a general willingness to describe oneself as falling somewhere on a social desirability continuum, rather than in terms of a simple acquiescence response set.

DISCUSSION

The results of these studies indicate that, for the attitude scales studied, items have definite social desirability values, that people respond differentially to these values, and that a substantial amount of variance in the attitude-scale scores can be explained as being due to this general response tendency. It is also apparent that the influence of the social desirability set affects scale items drawn from both personality and attitude scales. What do these findings mean in terms of scale construction and scale interpretation?

The answer to this question would seem to depend in part on the view one has on attitude scaling in general. If one defines attitudes solely in operational terms as being "what the attitude scale measures," it is apparent that the attitude scales studied here all to some extent measure a willingness to describe oneself in socially undesirable terms. From this point of view many attitude scales may be assumed to measure the general social desirability set, as well as measuring other more specific attitudes. If scales are wanted that measure specific attitudinal factors free from conflicting response sets, then one would be advised to cancel out the influence of social desirability (as Eysenck seems to have done with some success in his Radicalism scale). In any event, a certain caution seems called for in the labeling and interpretation of present attitude scales. Many interscale correlations have probably been reported as proving some quite specific hypothesis, when in fact such correlations have resulted from general social desirability

It is, of course, possible to conceive of specific attitudes as being much more than "what the attitude scale measures." Perhaps the more usual view is that attitudes can better be thought of as "a readiness to respond in certain ways," a readiness that may show itself in verbal behavior, overt action, fantasy, tension states, etc. From this point of view the social desirability variable can perhaps best be conceived as rather overriding attitude, a single attitude that interacts with and sometimes masks other attitudes. For instance, "latent anti-Semitism" may be regarded as an anti-Semitic attitude that is masked by the social desirability variable. If so, the researcher may be advised to take social desirability and its effects into account when he studies the influence of other specific attitudes on behavior.

The notion of a general social desirability set—influencing attitude scales and personality inventories alike—raises a further theoretical question: What factors lie behind differences in the set? On a priori grounds we might wonder whether it is influenced by such variables as "conformity," "psychopathology," "low status," "poor morale," or "low self-esteem." The reader can probably add to this list ad infinitum. Although the concept of social desirability serves to clarify certain problems in attitude and personality scales, it also adds a further measure of complexity to the researcher's task.

SUMMARY

This study investigates the influence of social desirability on attitude items and on attitude scale scores. Items were drawn from six attitude scales: the California Authoritarian Scale, the California Ethnocentrism Scale, the Srole Anomie Scale, the Eysenck Radicalism Scale, the PARI Approval of Hostility scale, and the PARI Approval of Control scale. It was found that definite social desirability norms exist for attitude-scale items; the median social desirability value of an item is highly correlated with the probability that the item will be endorsed; people differ in the extent to which they will endorse socially undesirable attitude statements, and this difference in willingness is a general response tendency which occurs irrespective of

the obvious content of the scale; a substantial correlation exists between this general response tendency and scores on five of the six scales investigated; and the tendency to describe oneself as undesirable on attitude-scale items is substantially correlated with the tendency to describe oneself as undesirable on personality-scale items. Data are presented to show that these results are not explainable by reference to a general acquiescence response set. It is possible that many interscale correlations have been reported as proving a specific hypothesis, when in fact such correlations have resulted from the more general social desirability variable. These findings sound a cautionary note in the development and interpretation of attitude scales, while at the same time raising certain theoretical problems.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality.* New York: Harper, 1950.

EDWARDS, A. L. *The social desirability variable in personality assessment and research.* New York: Dryden, 1957.

EYSENCK, H. J. *The psychology of politics.* New York: Praeger, 1955.

LENTZ, T. F. Acquiescence as a factor in the measurement of personality. *Psychol. Bull.*, 1938, **35**, 659. (Abstract)

SCHAEFER, E. S. Organization of maternal behavior and attitudes within a two-dimensional space: An application of Guttman's Radex theory. Paper presented at annual meeting of the Amer. Psychol. Ass., 1957.

SCHAEFER, E. S., & BELL, R. Q. Development of a Parental Attitude Research Instrument. *Child Develpm.*, 1958, **29**, 339–361.

SROLE, L. Social integration and certain corollaries: An exploratory study. *Amer. sociol. Rev.*, 1956, **21**, 709–716.

TAYLOR, J. B. Social desirability and MMPI performance: The individual case. *J. consult. Psychol.*, 1959, **23**, 514–516.

(Received March 23, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 391–400

THE INTERPRETATION OF PRAISE AND CRITICISM AS A FUNCTION OF THEIR SOCIAL CONTEXT

MORTON DEUTSCH[1]

Bell Telephone Laboratories, Murray Hill, New Jersey

In a previously published study (Deutsch & Solomon, 1959) it was found that the way a person reacts to a positive or negative evaluation of his performance is influenced by his own evaluation of the performance. The individual ("ego") tends to react more favorably to the other person ("alter") when alter's evaluation is consistent, rather than inconsistent, with ego's self-evaluation.

In a subsequent study (Deutsch, 1960) we asked observers to predict ego's reaction to alter's evaluation. We found that observers made systematic errors which were termed the "pathetic fallacy." Observers' evaluations, as compared to those of the actual participants, were determined not by the consistency between alter's evaluation of ego and ego's self-evaluation, but rather by the degree to which alter's behavior was consistent with perceived social norms concerned with the "maintenance of face." That is, the more alter went out of his way to help ego maintain face, the more favorably was alter evaluated. Similarly, the more alter went out of his way to destroy ego's face, the less favorably was alter evaluated.

The foregoing studies are part of a program of research concerned with understanding the factors that influence the interpretation of praise and criticism. We define "praise" to mean any positive evaluation of ego by alter and "criticism" to mean any negative evaluation of ego by alter. We shall not attempt here to develop a system for the description and classification of evaluations (see Kluckhohn, 1951, for an attempt at such a system) but it is well to recognize that evaluations differ not only in sign (i.e., positive or negative) but also in such dimensions as: intensity, scope (e.g., are they applied to all of ego or to some limited aspect of ego such as a performance, a membership, a belief, a possession?), explicitness, type of value appealed to in the evaluation, mode of expression of the evaluation, to whom the evaluation is expressed, etc.

In the research reported in this paper, we have employed simple evaluative acts: "You did that extremely well," "You did that well," "You did that poorly," and "You did that extremely poorly." We have placed the different evaluative acts into various institutional settings: family, school, military, and work contexts. Within each context we have selected the three role relations defined by the relative power of alter as compared to ego: alter is superordinate to ego, alter is subordinate to ego, and alter and ego are peers. We are interested in determining how the attributes of an evaluative act in conjunction with the institutional and role context of the act influence its perception or interpretation.[2] We focus, in this paper, upon the interpretations of an *observer* rather than upon the interpretations of the participants in the interaction. We do *not* make the assumption that the interpretations of the observers are the same as those of the participants; our prior research (Deutsch, 1960) has indicated to the contrary. The interpretations of the actual participants are bound to be more complexly determined than those of detached observers. The participants' interpretations reflect not only the impersonal characteristics of the social situation, but also their self-attitudes in the situation and the particularistic aspects of the existing relationship with the other.

The basic assumption underlying the present research is that the framework for the perception of any social act, by someone not self-involved in the situation, is determined by the perceived social norms regulating the social situation in which the act occurs. More specifically, we assume that *there is a tendency to assimilate or to contrast the act to the social norms regulating the social situation in which the act occurs: assimilation occurring when the*

[1] I wish to acknowledge the assistance of Robert M. Krauss in the collection and analysis of the data. I am also indebted to M. B. Wilk for various suggestions concerning statistical procedures.

[2] Throughout the paper we shall use the words "perception," "cognition," and "interpretation" as synonyms.

differences between the act and its social context are small; contrast occurring when the differences are large. Thus, we assume that the interpreter attempts to arrive at an interpretation of the act and of its consequences that is stable and consistent with his views of the social norms within the institutional and role setting in which the act has occurred. Further, we can expect a tendency to perceive the various components of a social act (e.g., its content, its tone, its style) homogeneously as assimilated or as contrasted to the social context in which it occurs (Heider, 1958).

Moreover, we assume that the stable and coherent perception of an act of social deviance requires a more complex, articulated organization of the social elements being perceived (i.e., the act, ego, alter, the role relations, the institutional context) than does the stable and coherent perception of an act of social conformity. In everyday terms, one does not have to invoke special principles to understand why a person conforms to the social norms but one does to understand why a person violates them.[3] Hence, it is easier to arrive at a stable, coherent perception of conforming behavior.

Thus, one can expect a bias toward perceiving a social act as conforming or appropriate to the social norms of the context in which it occurs (i.e., *assimilating* the act to its context) rather than toward perceiving it as a violation of the social norms (i.e., *contrasting* the act to its context). Presumably, acts conforming to the social norms are more positively evaluated than acts violating the social norms by a participant in the social situation within which the act occurs (i.e., by someone who may be expected to accept the social norms).

We assume that social norms exist within American society which define the degree of appropriateness or inappropriateness of given acts of praise or criticism and we also assume that the social norms are somewhat different in different institutional and role contexts. Further, for our experiment, we assume that our subjects knew the norms pertaining to the different institutional and role contexts and that their interpretation of an act of praise or criticism was determined by the perception of the act in relation to the relevant social norms. Our initial assumptions concerning the relations between the social norms and the three independent variables of our study (i.e., "evaluative statement," "role," and "institution") were: the norms of the American culture are biased toward viewing praise as more appropriate than criticism; the norms with regard to authority are such as to define evaluations from a superordinate as more appropriate than evaluations from a subordinate; and the norms governing interactions within the family are the most permissive and benevolent, while the norms within the military are the least permissive and benevolent of the norms characterizing the four institutional settings considered in this study. Our more specific assumptions about the interactions among our independent variables will be discussed in connection with the results.

[3] The assumption that conformity requires no special explanation while deviance does is similar to a basic assumption involved in Parson's (1951) theoretical system: " . . . The continuance of a stabilized motivational process in a stabilized relationship to the relevant objects is to be treated as *not problematical* . . . . Hence for the social system as well as the personality we will *not* be concerned with the problem of the maintenance of given states of the social system *except when there are known tendencies to alter those states*" (p. 204).

METHOD

The study was conducted with students at several local colleges during a regular meeting of their psychology classes. By a random procedure, the students were assigned to the 48 experimental variations that resulted from our $4 \times 4 \times 3$ factorial experimental design. Sixteen students were assigned to each experimental variation; each student was exposed to five experimental variations.

The method of the study was to present the students (i.e., the observers) with simple evaluative statements made by alter to ego in which the role and institutional contexts were specified by appropriate labeling of alter and ego. The method is one that we arrived at after rejecting such methods as films, voice recordings, and puppet enactments for depicting the interactions. It became obvious that a simulation procedure runs into the problem that in any lifelike interaction the participants use a variety of techniques to define the meaning they are intending to convey by a given act. Thus, the meaning of a given act may be determined by such factors as the time and place in which it occurs, the tone of voice, gestures, manner, and appearance with which it is accompanied, etc. In real life, one attempts to manipulate these expressive components of an act to produce in a given social context, with its given social norms, the effect one intends. Thus, a private who wants to encourage a captain through praise will express his praise in a tone of voice, with gestures and a manner, etc. that are different from the tone of voice, etc. of a captain who wants to encourage

a private. Hence, the experimental simulation of praise would have to be rather different for the different roles and social contexts if one is to avoid introducing a false note or inconsistency into the act itself. The radical simplification of our procedure[4] largely eliminates the expressive components of the act and, hence, allows one to study the effects of social context upon acts that are otherwise identical (and which, despite their identity, do not have false or inconsistent notes in any of the social contexts).

By varying the evaluative statement and the labeling, the following experimental variables were created:

Evaluation scale. Four evaluative statements were used to create four levels of this factor: "You did that extremely well," "You did that well," "You did that poorly," "You did that extremely poorly."

Institutional context. Four institutional contexts were employed: family, school, work, and military. The institutional context was created by the labeling of ego and alter. For the family, the labels father, son, and brother were employed; for the school, teacher and student; for work, employer and employee; for the military, captain and private.

Role context. Three role contexts were created: alter as superordinate to ego (i.e., father to son, teacher to student, employer to employee, captain to private); alter and ego as peers (i.e., brother to brother, student to student, employee to employee, private to private); and alter as subordinate to ego (i.e., son to father, student to teacher, employee to employer, private to captain).

Examples of the statements are: "You did that extremely well," Mr. Pitman, a student, told his teacher, Mr. Anderson. "You did that poorly," Bob Simon said to his father. "You did that extremely poorly," Private Ford told Private Walker.

Data were obtained by having the observers answer four questions: What did ego (i.e., the recipient of the act) think alter's (i.e., the person making the evaluation) purpose was? What did ego say to alter? What did ego want to say to alter? and What did ego feel? The questions were, of course, phrased not in terms of ego and alter but in terms of the names of the people in the depicted interaction. The observers responded to the questions by filling in 13-point rating scales, which ranged in value from −6 through 0 to +6. The scales for Questions 1 and 4 were identical: +6 was labeled "extremely good," 0 was "neither good nor bad," −6 was "extremely bad." The instructions, of course, made it clear to the observers that for Question 1 a rating of +6, for example, meant that they (i.e., the observers) held the view that ego thought alter's purpose in making his statement was "extremely good," while a rating of +6 in response to Question 4 meant they thought ego felt "extremely good" after hearing alter's statement. The scale for Questions 2 and 3 were identical, +6 was labeled "extremely friendly," 0 was "neither friendly nor unfriendly," −6 was "extremely unfriendly." The observers understood that a rating, for example, of +6 on Question 2 meant they thought ego said something that was "extremely friendly" in reply to alter and that a rating of +6 on Question 3 indicated they thought ego wanted to say something "extremely friendly"; a difference in ratings on the two questions represented a difference between what they thought ego wanted to say and what they thought he actually said. In addition to the rating scale data, free responses were obtained to the four questions from 337 observers.

[4] In principle, the methodology of our study is no different from the methodology employed in the laboratory investigation of physical perception. There, too, radical simplification rather than simulation is the method of choice in studying basic processes of perception.

RESULTS

The main data for this experiment were obtained from the observers' ratings in response to the question: What did ego think alter's purpose was? Since the ratings for all four questions were highly intercorrelated (the average r was .71), the other three ratings are not discussed as such. They are, however, considered in relation to measures derived from them that are largely free of "halo effects" and "response bias." These measures, which are discussed more fully below, are "the perceived immediate emotional impact" of alter's evaluation upon ego and the "perceived behavioral restraint" of ego in responding to alter.

Observer's Evaluation of Alter's Purpose

General effects. Tables 1 and 2 provide a summary of the data concerning the observer's evaluation of alter's purpose. It is apparent that the same act (e.g., "you did that extremely well") is evaluated in strikingly different ways as a function of the context of the act (e.g., compare the evaluation when a father is saying it to his son with a private saying it to a captain). The effect of context can be so strong that intense criticism of a son by a father is more favorably evaluated than intense praise of a captain by a private.

If one accepts our assumption that a positive evaluation of alter's purpose indicates that the observer has perceived that alter is conforming to the social norms governing his relationship with ego and that a negative evaluation indicates the opposite, it is evident that, over all, there is a strong bias in favor of perceiving alter as conforming to rather than violating the norms. We have suggested that the bias results from the greater ease of arriving at a stable, coherent perception of con-

TABLE 1
CELL MEANS: EGO'S EVALUATION OF ALTER'S PURPOSE (as perceived by observers)

Scale	Status	Institution				$\bar{X}$ Scale
		Family	School	Work	Military	
Extremely well	high–low	4.94[a]	4.88	4.38	2.81	
	low–low	4.31	3.25	4.00	2.50	
	low–high	3.88	2.75	1.69	.81	
						(3.35)
Well	high–low	3.38	3.69	3.50	3.81	
	low–low	2.50	2.19	3.75	2.38	
	low–high	2.62	2.06	1.69	.38	
						(2.66)
Poorly	high–low	.06	1.12	1.00	.31	
	low–low	1.00	.94	−1.31	.44	
	low–high	.50	−.56	.56	−3.75	
						(.03)
Extremely poorly	high–low	1.88	−.12	.00	1.19	
	low–low	.31	−.62	−2.25	.06	
	low–high	−1.00	−2.31	−.25	−2.12	
						(−.44)
$\bar{X}$ Institution		(2.03)	(1.44)	(1.40)	(.73)	
$\bar{X}$ Status						
high–low	(2.30)					
low–low	(1.46)					
low–high	(.43)					

[a] Each cell entry is a mean based on 16 responses.

TABLE 2
ANALYSIS OF VARIANCE: EGO'S EVALUATION OF ALTER'S PURPOSE (as perceived by observers)

Source	*df*	*MS*	*F*
Between institution	3	53.95	6.48***
Between scale	3	681.85	81.95***
Between status	2	223.94	26.92***
Institution × Scale	9	13.65	1.64*
Institution × Status	6	22.33	2.68**
Scale × Status	6	3.18	0.38
Institution × Scale × Status	18	12.99	1.56*
Within cells	720	8.32	

* $p < .10$.
** $p < .05$.
*** $p < .01$.

forming as contrasted with deviating behavior.[b]

Effects of the evaluative statement. Tables 1 and 2 also give evidence of a bias that results in more favorable interpretations of praise than of criticism (the bias is intensified by more intense praise or criticism) even when the effects of role and institutional context are eliminated statistically. We use the term bias here to indicate that we can see no intrinsic reason why praise, per se, should be more favorably evaluated than criticism, per se. The finding, although intuitively obvious, requires an explanation. Our explanation assumes that the purpose of any social act is always evaluated in relation to some normative context; the context that remained for the observers, after the elimination of the specific effects of "role" and "institution," was the broad normative context of American society. Elsewhere (Deutsch, 1960), we have developed the rationale for assuming that the norms within the American society are biased in favor of praise. One can imagine other cultures that might be strongly biased in terms of "appropriateness of evaluation" or even some "masochistic" cultures that would be biased toward favorable evaluation of criticism.

Effects of role. We have assumed that the social norms concerning authority are such that they would lead the observers to regard alter's evaluative act more favorably the

[b] In another study, with the same social context variables, we presented an interaction in which alter's evaluative act clearly violated all of the relevant social norms: he was described as *ridiculing* ego. Ridicule is evaluated quite negatively in each of the institutional and role contexts. The contrast between the act of ridicule and the social norms is so strong as to "wash out" the differential effects of role and institutional contexts.

higher alter's status is in relation to ego. The results clearly support this assumption. For each of the four evaluative acts, the higher alter's relative status the more favorably do the observers interpret the purpose of the act. There is evidence (reflected in the "scale and status" interaction) that the effect of relative status upon the observers' interpretations is enhanced when the evaluation is intense rather than mild. That is, subordinates are perceived to have even relatively less right, as compared with superordinates, to make intense rather than mild evaluations. Status differences have most effect upon interpretations of acts within the military, next most in the school, and least in the family, (cf. the "institution and status" interaction). The latter results are consistent with the assumption that the relative status of alter to ego is a more important determinant of the observers' evaluations of alter's act, the more rigid the authority structure[6] within which the act occurs.

Effects of institution. We have assumed that the observers' interpretations of alter's act would be more favorable if alter's act takes place within an institutional context of relations between alter and ego that is perceived to be cooperative or mutually facilitative rather than individualistic or competitive. From this assumption, one would expect that an evaluative act would be most favorably interpreted in the family and least favorably in the military context. The results confirm this expectation.

Within the school context, evaluative acts between superordinate and subordinate tend to be interpreted more favorably than in the work context unless the evaluation is one of intense criticism, in which case the reverse is true. On the other hand, evaluations between peers are more favorably interpreted in the work than in the school context unless the evaluation is critical, in which case the reverse is true. These reversals indicate a more intense contrast effect for negative evaluation of peers in the work as compared to the other settings and a more intense contrast effect for severe criticism of a person of different status in the school as compared to the other contexts.

As a possible explanation of these results we offer the hypothesis that a contrast effect is likely to occur more intensely or with weaker differences, the more it is perceived possible for relations to occur between alter and ego opposite to those that usually obtain. Thus, observers might assume that within the family context there is little possibility for a relationship that is not mutually facilitative to occur. Here it would take a large violation of the social norms (see Footnote 5) to produce a contrast effect. On the other hand, between teacher and student or between co-workers, observers might assume that while a mutually facilitative relationship is most common, a mutually antagonistic relationship is not uncommon. Here, a violation of the norms would more readily lead to a contrast effect. This hypothesis explains the findings described in the previous paragraph if one assumes that observers are more likely to perceive relations between different pairs of employers and employees as more homogeneous than among different pairs of co-workers, and co-student relationships as more homogeneous than teacher-student relationships. We have no evidence that is relevant to this assumption.

Perceived Immediate Emotional Impact upon Ego of Alter's Evaluation

From the difference between the observers' ratings of "what ego felt" and "what ego thought alter's purpose was" a measure of the perceived immediate emotional impact of alter's evaluation upon ego was obtained. It seems reasonable to assume that the total emotional impact of an act (i.e., "what ego felt") results from its perceived purpose and its immediate emotional impact. The immediate emotional impact of an evaluation reflects, we believe, its impact upon the individual's self-esteem (or upon his security system). That is, an evaluative act has emotional impact for a person over and above the significance attributable to its perceived purpose if the evaluation is perceived to have relevance to the

[6] "Rigidity of authority" is difficult to define in terms of a single criterion. However, such indicators as the following would reflect this concept: the right of a superordinate to expect unquestioning obedience from a subordinate; the severity of the sanctions that a superordinate can legitimately invoke to punish a disobedient subordinate; the freedom of the subordinate to appeal to others to countermand an order from his superordinate or to support his violation of an order.

TABLE 3
CELL MEANS: IMMEDIATE EMOTIONAL IMPACT OF ALTER'S EVALUATION UPON EGO
(as perceived by observers)[a]

Scale	Status	Institution				$\bar{X}$ Scale
		Family	School	Work	Military	
Extremely well	high–low	.38	.56	.38	.06	
	low–low	.88	.12	−.38	1.25	
	low–high	.31	.56	.50	1.81	
						(.44)
Well	high–low	.44	.38	1.06	1.06	
	low–low	.38	1.19	.88	1.00	
	low–high	.38	.88	.06	1.69	
						(.78)
Poorly	high–low	2.44	3.00	2.62	1.88	
	low–low	1.88	2.31	.81	.06	
	low–high	2.25	1.69	1.31	−.81	
						(1.62)
Extremely poorly	high–low	3.69	2.19	2.75	3.38	
	low–low	2.00	1.12	.12	.50	
	low–high	1.31	.25	1.19	1.19	
						(1.62)
$\bar{X}$ Institution		(1.36)	(1.09)	(.92)	(1.09)	
$\bar{X}$ Status						
high–low	(1.64)					
low–low	(.84)					
low–high	(.87)					

[a] Cell entries for this table were obtained in the following manner: for positive evaluative statements ("extremely well" and "well") the difference was computed as "ego felt" minus "alter's purpose." For negative evaluative statements ("poorly" and "extremely poorly") the difference was computed as "alter's purpose" minus "ego felt."

TABLE 4
ANALYSIS OF VARIANCE: IMMEDIATE EMOTIONAL IMPACT OF ALTER'S EVALUATION UPON EGO
(as perceived by observers)

Source	*df*	*MS*	*F*
Between institution	3	6.28	.90
Between scale	3	68.67	9.85***
Between status	2	52.92	7.59***
Institution × Scale	9	20.25	2.90***
Institution × Status	6	5.41	.78
Scale × Status	6	25.26	3.62***
Institution × Scale × Status	18	5.34	.77
Within cells	720	6.97	

*** $p < .01$.

values and social supports upon which the person bases his self-esteem.

The data (summarized in Tables 3 and 4) indicate that criticism is perceived to have more immediate impact than praise. Criticism from a superordinate is perceived to have more immediate impact than criticism from a peer or a subordinate. However, there is a suggestion in the data that for praise the reverse is true in the military. These findings seem consistent with our view that immediate emotional impact reflects self-esteem involvement. Here, we assume as we have elsewhere (Deutsch, 1960) that in a success-oriented culture such as ours, threat to self-esteem (as in failure or criticism) is perceived to be more self-involving than is reassurance (as in success or praise). It is also reasonable to assume that the threat from criticism is perceived to be greatest when the critic is a superordinate.

Initially, we had also assumed that the personal-impersonal characteristics of an institution contribute to the perceived emotional impact of an evaluative act and, hence, that the immediate emotional impact would be perceived to be greatest in the family. However, this assumption does not provide an adequate explanation for praise having more marked impact in the military than in the family context. Alternatively, one may hypothesize that an evaluative act has immediate emotional impact, not only as a function of its relevance to the individual's security system, but also as a function of its surprise or its unexpectedness. Thus, praise in the military (as compared with the family), especially from a subordinate, and criticism in the family (as compared with the military) are both likely to be considered relatively unexpected.

The perceived behavioral restraint of ego in responding to alter. The data presented so far have been concerned with the observer's expectation concerning the effect upon ego of alter's evaluation of ego. Our interest here was in the observer's view of ego as the *recipient* of an evaluation. It is also relevant to consider, from the observer's viewpoint, ego as a *responder* to alter's evaluation. In general, it is reasonable to posit that the observers would assume a large measure of complementarity in the social norms that determine ego's reception or interpretation of an evaluation and the norms that determine ego's action toward alter in response to alter's evaluation. Our research design, however, does not permit any unequivocal test of this assumption. A correlation of .76 ($p < .01$) is obtained between the observers ratings of alter's purpose and of ego's overt response to alter; the correlation in addition to reflecting the assumed complementarity undoubtedly is also influenced by "halo effects."

While complementarity is to be expected, there is also reason for the observers to assume that ego's overt behavior toward alter may not necessarily be the same as his covert behavior; that social norms parallel to the ones affecting his interpretation of alter's behavior would place restraints upon ego's overt response. These perceived restraints can be measured by comparing overt and covert responses. *Negative restraint* may be defined as existing whenever the overt response is perceived to be more favorable than the covert response toward alter; *positive restraint* as existing whenever the covert response is perceived to be more favorable than the overt response. Our initial assumptions focused upon the occurrence of negative restraint. We anticipated that the observers would expect more negative restraint after criticism when ego was a subordinate rather than a superordinate of alter and when the institutional setting was more rigidly authoritarian and less benevolent (e.g., the military as compared with the family). We also expected more negative restraint after criticism than praise.

The data (see Tables 5 and 6) show there to be more negative restraint in response to criticism than to praise; in certain social settings ego is expected to show positive restraint in relation to praise. Thus, in response to praise from peers or from members of his family, ego is expected to make an overt response that is less positive than his covert response. This suggests that in certain situations a restraint is perceived against *ingratiating behavior*, that may be sufficiently strong to curb the overt expression of gratitude. One may speculate that such a curb arises as a means of protecting or insuring the sincerity of praise when, frequently, in other social contexts it is an attempt at ingratiation.

The expectation of greatest perceived negative restraint after criticism from a superordinate is supported by the data. However, the institutional effects were not as anticipated. In the military, where one would expect the largest amount of perceived negative restraint after criticism by a superordinate, one finds the smallest. One possible explanation for this surprising result is that the negative restraint in responding to criticism from a superordinate in the military are perceived to be so great that they are perceived to inhibit covert as well as overt action toward alter. This interpretation is given some plausibility by other data (not presented here) which indicate that a measure of perceived *internal restraint* (the difference between the observers' ratings of "what ego felt" and "what ego wanted to say") is highest in the military after criticism from a superordinate.

Analysis of the Free Response Data

The data collected from the observers when they gave free responses to the four questions (concerning alter's purpose, ego's overt and covert responses to alter, and ego's feelings) are presented as supplementary, secondary data because of the methodological difficulties we experienced with them. About 25% of the observers' responses were uncodable and it was impossible to code many of the codable responses without knowledge of the experimental treatments since many of the responses specifically identified the institutional and, or, role contexts. Thus, we are unable to rule out the possibility that the analysis of these data is biased by knowledge of the experimental treatments. Nevertheless, since they give results completely consistent with the rating scale data, it seems worthwhile to take advantage of the insights they offer.

The perception of alter's purpose. Praise, as

TABLE 5

CELL MEANS: DIFFERENCE BETWEEN WHAT EGO SAID AND WHAT HE WANTED TO SAY TO ALTER (as perceived by observers)[a]

Scale	Status	Institution				$\bar{X}$ scale
		Family	School	Work	Military	
Extremely well	high–low	.25	.06	.25	.19	
	low–low	−1.00	−.44	−.69	−.19	
	low–high	.25	.56	.69	−.12	
						(−.06)
Well	high–low	−.31	.00	.06	−.12	
	low–low	−.81	1.44	−.62	−.06	
	low–high	.44	.12	1.06	.56	
						(.07)
Poorly	high–low	.88	.69	.94	.38	
	low–low	.06	.62	.94	.94	
	low–high	.50	1.69	.19	−.06	
						(.65)
Extremely poorly	high–low	.62	1.44	1.81	.25	
	low–low	−1.06	.69	.69	1.06	
	low–high	−.44	−.62	.75	.19	
						(.45)
$\bar{X}$ Institution		(−.17)	(.52)	(.51)	(.25)	
$\bar{X}$ Status						
high–low	(.43)					
low–low	(.10)					
low–high	(.30)					

[a] The cell entries in this table were computed by subtracting the observers' responses on "ego said to self" from "ego said to alter."

TABLE 6

ANALYSIS OF VARIANCE: DIFFERENCE BETWEEN WHAT EGO SAID AND WHAT HE WANTED TO SAY TO ALTER (as perceived by observers)

Source	*df*	*MS*	*F*
Between institution	3	19.78	3.34
Between scale	3	20.39	3.44**
Between status	2	7.20	1.22
Institution × Scale	9	3.21	0.54
Institution × Status	6	6.23	1.05
Scale × Status	6	9.74	1.65
Institution × Scale × Status	18	4.83	0.82
Within cells	720	5.92	

** $p < .05$.

contrasted with criticism, is much more frequently perceived to be motivated by "ought" forces. That is, praise is seen as being motivated by a desire to be "truthful" or "objective" more than twice as frequently as criticism. The evaluations made by superordinates are more likely to be perceived as motivated by ought forces than are the evaluations by peers or by subordinates. Praise by a peer is most frequently perceived as motivated by a desire to make ego "feel good" or "happy," praise by a superordinate is most frequently seen as having the intention of "helping" ego (e.g., "letting him know that he is right," "encouraging him," "teaching him"), while praise from a subordinate is most commonly interpreted as being motivated by the subordinate's own needs (e.g., "to ingratiate himself with his superior," "to get the boss to do something," "to be a big shot"). The differences in interpretation of the purpose of praise for the different status relations is most accentuated for the military setting and virtually disappears in the family setting. In the latter setting, praise by a brother or by a son are both seen, most often, as a desire to make ego "feel good." In the more impersonal institutional settings, praise is more likely to be seen as intending to be "helpful" than as intending to make ego "feel good"; the opposite is true for the family setting.

Criticism by superordinates, especially in the family setting, is seen typically as intending to be helpful. In all but the work setting, criticism by peers is also perceived as intending to be helpful (although less frequently than in the case of criticism by superordinates); in the work setting, it is perceived as malicious in intent (i.e., making ego "feel bad"). Criticism by subordinates is most often viewed as negative in purpose in the military and school contexts but not in the family or work con-

texts. The more intense the criticism, the more likely that the subordinate's purpose is described as malicious rather than as an attempt to be a "big shot."

Ego's affective reaction to alter. In all of the institutional settings, praise is perceived as giving rise to positive affect which is primarily self-directed (e.g., "feeling of self-esteem," "feeling of achievement") rather than alter-directed (e.g., "feeling of solidarity," "feeling of respect"). Criticism in the school and family settings is also seen as giving rise to self-directed feelings (e.g., "feeling of insecurity," "feeling of failure"); but, in the military and work settings, criticism gives rise to alter-directed affect (e.g., "contempt or pity," "anger or indignation"). Praise or criticism, but especially criticism, from a subordinate is more likely to be perceived as evoking alter-directed affect than evaluations from a peer or a superordinate.

The affective reaction of contempt, pity, anger, or indignation to criticism is expected most often when alter is lower in status than ego and least frequently when alter is of higher status. "Resentment" is expected in reaction to criticism only when alter is of higher status than ego, and only in the military and work contexts. The feeling of determination to improve one's performance is expected most often after criticism by a superordinate and least often after criticism by a subordinate.

DISCUSSION

The data of this experiment clearly demonstrate that the meaning of praise and criticism is very much influenced by the social contexts in which they occur. The data are consistent with everyday observations and with common sense intuitions even though they may run counter to the theoretical notions behind such concepts as "generalized reinforcers" (Krasner, 1958; Skinner, 1957). On theoretical examination (Deutsch & Solomon, 1959), the obvious fact that "praise" is, generally, more positively responded to than "criticism" no longer remains obvious; it becomes a fact requiring explanation. Our explanation has been that a social act is always interpreted in relation to some normative context(s). For the disinterested observer of a social act, we have assumed that the relevant normative framework within which an act is perceived includes the perceived, relevant norms from the culture, the institutional setting, and the role relationship within which the act occurs. For the actual participant in a social act, we assume that norms which relate to his self-conception and to the particularistic relations with the other also influence the framework within which an act is perceived.

Praise tends to be more favorably evaluated than criticism, in terms of this explanation, because the norms of the success-oriented American culture are directed toward helping people maintain the "face" of success (Kluckhohn, 1949; Williams, 1951) and because acts that are perceived as conforming to the accepted norms are evaluated more favorably than acts that violate them. However, as this and a prior experiment (Deutsch & Solomon, 1959) have demonstrated, when praise is counter to and criticism is consistent with specific institutional, role, or self-norms the reverse evaluations can be obtained.

Despite the seeming superficiality and artificiality of the methodology we have employed, we believe that our results indicate that the technique is a very useful method for the investigation of social norms. The technique is, of course, similar to certain projective methods (e.g., the sentence completion method) used to study individual personality. Here, however, the emphasis is on social situational variables. By systematically varying the social context of standard social acts and obtaining both quantitative and qualitative responses, one should be able to characterize in comparable terms the perceived effects of age, sex, occupational and other characteristics upon the meaning of a given social interaction.

SUMMARY AND CONCLUSIONS

In this paper, we have investigated the effects of the institutional and role context upon the interpretation by a disinterested observer of written presentations of simple acts of praise and criticism made by alter to ego. We placed the different evaluative acts (e.g., "You did that extremely well," "You did that extremely poorly") in different institutional contexts (family, school, work, and military), and within each context we selected three role relations defined by the relative power of alter as compared with ego: alter is superordinate to ego, alter is subordinate to ego, and

alter and ego are peers. From the observers, we collected ratings or free responses to four questions: What did ego (i.e., the recipient of the act) think alter's (i.e., the person making the evaluation) purpose was? What did ego say to alter? What did ego want to say to alter? and What did ego feel?

The results may be summarized as follows (keeping in mind that the data refer to an observer's predictions concerning ego's reactions to alter's evaluations of ego):

1. There is a bias toward interpreting evaluative acts favorably, i.e., toward viewing them as conforming to rather than violating the social norms.

2. There is a bias toward interpreting praise more favorably than criticism. Criticism is perceived as having greater immediate emotional impact than praise and as being more likely than praise to result in "holding one's tongue."

3. Evaluations made by a superordinate to a subordinate are interpreted most favorably, evaluations by a peer next most favorably, evaluations by a subordinate least favorably. The immediate emotional impact of evaluations by superordinates is expected to be greater than those made by subordinates. The "holding of tongue" in response to intense criticism is expected to be greatest by subordinates and least by superordinates.

4. Evaluative acts are interpreted most favorably within the family, somewhat less favorably in school and work, and least favorably within the military context. The effects of status differences upon the interpretation of an evaluative act are seen to be most marked within the military and least marked within the family. The immediate emotional impact of criticism is expected to be greatest in the family, the immediate emotional impact of praise is expected to be greatest in the military.

REFERENCES

DEUTSCH, M. The pathetic fallacy: An observer error in social perception. *J. Pers.*, 1960, **22**, 317–332.

DEUTSCH, M., & SOLOMON, L. Reactions to evaluations by others as influenced by self-evaluations. *Sociometry*, 1959, **22**, 93–112.

HEIDER, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

KLUCKHOHN, C. *Mirror for man.* New York: Whittlesey House, 1949.

KLUCKHOHN, C. Values and value-orientations. In T. Parsons & E. Shils (Eds.), *Toward a general theory of action.* Cambridge: Harvard Univer. Press, 1951. Pp. 359–433.

KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148–170.

PARSONS, T. *The social system.* Glencoe, Ill.: Free Press, 1951.

SKINNER, B. F. *Verbal behavior.* New York: Appleton-Century-Crofts, 1957.

WILLIAMS, R. M., JR. *American society.* New York: A. A. Knopf, 1951.

(Received March 24, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 401–407

QUALITY AND ACCEPTANCE OF PROBLEM SOLUTIONS BY MEMBERS OF HOMOGENEOUS AND HETEROGENEOUS GROUPS[1]

L. RICHARD HOFFMAN AND NORMAN R. F. MAIER
University of Michigan

In an earlier article Hoffman (1959) reported that higher quality solutions to problems were produced by groups whose members had dissimilar personalities (heterogeneous groups) than by more homogeneous groups. On the other hand, the reported satisfaction of the members with the solutions (acceptance) was quite similar in the two types of groups. Since only two problems were used, however, the generality of the superior problem solving ability shown by the heterogeneous groups may be questioned.

Might homogeneous groups do better on other problems? Problems that emphasize differences among group members, and thus strain the ability of heterogeneous groups to work together effectively, may provide conditions advantageous for homogeneous groups. Even if the quality of solutions to such problems by heterogeneous groups should not be impaired, the members' satisfaction with the solutions may be reduced.

In addition to the two problems used in the original study, the present study employed three new ones that involved conflicts in values, personal conflicts, and the distribution of a bonus. With the same experimental design the original study was thus extended to determine whether heterogeneous groups would produce higher quality solutions to such varied problems without sacrificing acceptance.

METHOD

Four-person homogeneous and heterogeneous groups were constructed in the eight laboratory sections during two semesters of an undergraduate psychology course in human relations. The Guilford-Zimmerman Temperament Survey (Guilford & Zimmerman, 1949) was used as the measure of personality, and profile correlations (Kendall's tau) were computed between the 10-score profiles of all pairs of students in each section. Originally 16 homogeneous groups (high positive tau's) and 25 heterogeneous groups (either near-zero tau's or high negative and positive tau's) were formed in the third week of the semester and interacted weekly thereafter in case discussions, problem solving, and role playing.

Periodically, test problems were introduced into the laboratory sessions. Solutions to four of these problems could be scored for their objective quality; one was a problem purely of fairness, having no external standard of quality. Of the problems that could be scored for quality, solutions to three (Mined Road, Student Assistance Fund, and Painter-Inspector Argument) were graded quantitatively according to systems of content analysis that assigned increasing numbers of points to solutions of higher quality. On the fourth quality problem—Change of Work Procedure—new-modified and, especially, integrative solutions were considered to be of high quality.

The Mined Road Problem, used in the earlier study, (Hoffman, 1959) requires the group to develop a method for permitting five men to cross a heavily mined road. It was administered in the first meeting of each group. The major problem for the group here is to recognize and solve all facets of the problem in the safest possible way.

The Student Assistance Fund problem, a role playing case developed for the present study, was administered in the fourth meeting. Each subject was assigned the role of a different student who was deemed qualified for financial assistance from a university fund. Each "student" was told that he qualified on one or more attributes. Each also was told that he needed $1,500 for the next academic year, but the total fund was only to be $3,000 this year. The group had to decide on how to allocate the money this year and to recommend a method for allocating money in the future. Conflict was induced by the provision that the total amount of money was less than that needed by all group members combined and not enough for any one member if they divided the money evenly. The chance of reaching agreement was reduced further by giving the subjects opposing standards for allocating money among themselves. Only the method developed for allocating funds among students in the future was scored for quality.

The Painter-Inspector Argument problem (adapted from Maier, Solem, & Maier, 1957), another role playing case in which a leader was assigned randomly within each group, was administered in the sixth laboratory session. A painter, an inspector and rough primer, a shop steward, and a foreman meet to settle

[1] The major part of the research reported herein was supported by Project Michigan, a Department of the Army-sponsored project of the University of Michigan in the field of combat surveillance. The contract (DA-36-039 SC 52654) is administered by the Signal Corps. The final report was written in conjunction with USPHS Grant No. M-2704 from the National Institute of Mental Health, United States Public Health Service. A shortened version of this paper was presented at the Midwestern Psychological Association meetings, April 1960.

an argument between the painter and the inspector. The argument reflects a basic conflict in the entire shop, but few groups were able to get beyond the initial argument and identify the total shop problem.

The Change of Work Procedure problem (Maier, 1952), administered in somewhat modified form in the eighth session, is also a role playing case, with a foreman appointed to lead the discussion, and was repeated from the earlier study. The discussion centers around a request by the foreman that each of the three workers work in his best position rather than rotating among three simple jobs. The problem is usually approached as a choice between the increased pay of the fixed-position method (New solution) and the relief from boredom offered by rotation (Old solution).

The Point Distribution problem was administered in the tenth session, immediately following the return to the class members of the third of four course examinations. Each group was given 19 points to distribute among its members as a bonus for participating in the experiment. The points were to be used to raise the final grade of the student. Each subject knew his own position in the class relative to the tentative cutting points for grades indicated by the instructor. The groups were given 40 minutes to arrive at a decision, and were told that no fractional or negative points were to be given and that failure to agree would give each subject four points. Since no objective standard of quality could be applied to the solutions, this problem was one of fairness or acceptance.

Following each problem solving session, the subjects completed two Likert-type five-point questions concerning their satisfaction with the solution reached and with the amount of influence they had over the solution. Values from 1 to 5 were assigned to each response, with 1 indicating "very dissatisfied" and 5, "very satisfied."

RESULTS

Quality of Solutions

The overall findings of the present study support those of the earlier one. On the various problems, the solutions by heterogeneous groups either were scored as significantly superior in quality to those of the homogeneous groups, or did not differ in quality. Never were the homogeneous groups better than the heterogeneous groups.

Table 1 presents a comparison of the results on the three quantitatively-scored problems that had a quality component—Mined Road, Student Assistance Fund, and Painter and Inspector—for the homogeneous and heterogeneous groups. For the Mined Road problem the results combining the data of the present and earlier studies are also presented. The proportion of high quality solutions to both the Mined Road and Student Assistance Fund problems by heterogeneous groups was substantially greater than by homogeneous groups. (The means are not very different as a result of a few very high scoring solutions by homogeneous groups.) The Painter and Inspector problem was the only one in which no differences were found between the two types of groups. The generally low scores on this problem (an average of about 15 out of a possible 45 points with less than 20% of the groups achieving scores of 30 or more) point to the restrictive character of this problem. Even heterogeneity of personality was insufficient to force the groups to consider other than the immediate problem of settling the personal dispute.

TABLE 1
QUALITY OF SOLUTIONS BY HOMOGENEOUS AND HETEROGENEOUS GROUPS

	Present Study		Combined Studies	
	Homogeneous	Heterogeneous	Homogeneous	Heterogeneous
Mined Road Problem				
Percentage above median[a]	30%	57%	21.7%*	52.6%*
M	55.9	65.3	45.9*	64.3*
SD	34.6	29.9	32.1	29.3
Number of groups	10	21	23	38
Student Assistance Fund				
Percentage above median[a]	36%*	80%*		
M	44.2	50.9		
SD	17.5	12.8		
Number of groups	14	20		
Painters and Inspectors				
Percentage above median[a]	54%	56%		
M	16.5	15.1		
SD	10.5	13.9		
Number of groups	13	16		

[a] The median score was determined from the results of all four-person groups, including residual groups not assigned to either the homogeneous or heterogeneous types.

* Percentage or mean difference significant at the .05 level.

The results in Table 2 for the Change of Work Procedure problem again show the superiority of the heterogeneous over the homogeneous groups. Whereas 65% of the heterogeneous groups in the present study produced new-modified or integrative solutions—the higher quality solutions—only 21% of the homogeneous groups' solutions were of these types. This difference is statistically significant at the .05 level. Furthermore, the percentage of integrative solutions—the highest quality solution—produced by the hetero-

TABLE 2

SOLUTIONS TO THE CHANGE OF WORK PROCEDURE PROBLEM BY HOMOGENEOUS AND HETEROGENEOUS GROUPS

Type of Solution	Homogeneous	Heterogeneous
Integrative	0 (7.7%) [a]	5 (35.3%) [a]
New-Modified	3*	6*
New	7	4
Old	4	2

[a] The numbers in parentheses are the percentages of Integrative solutions obtained from each type of group in the combined studies. This percentage difference is statistically significant at the .05 level.

* Chi square test comparing Integrative and New-Modified with Old and New solutions is significant at the .05 level.

geneous groups in both the present and earlier studies was significantly greater than the percentage produced by homogeneous groups (indicated in parentheses in Table 2).

Thus on three of the four problems that could be scored for quality, substantially more of the heterogeneous than of the homogeneous groups produced better solutions. On the fourth problem there was no difference between the group types, most solutions being of poor quality.

The Point Distribution problem was designed to place strain on the heterogeneous groups by exaggerating the differences among the group members. The problem demands agreement on the criteria to be used in assigning points to the several group members. Greater conflict was anticipated in the heterogeneous than in the homogeneous groups on the grounds that in the former groups the differing personalities would use different standards to determine the appropriate method of awarding points. Thus the heterogeneous groups were expected either to allocate the points nearly equally, reflecting an inability to resolve the conflict, or to distribute the points widely, reflecting some agreed-upon systematic method of awarding the points.

The data tend to support this prediction. Since all groups arrived at some conclusion, the standard deviation of points awarded in a group was computed and used as the measure of outcome. Small scores represent nearly equal distributions of points and large scores represent diverse distributions. Although the mean standard deviation scores for the two types of groups were approximately equal (3.06 for the homogenous groups and 4.08 for the heterogeneous groups), the standard deviation of these scores for the heterogeneous groups was significantly greater than for the homogeneous groups (2.95 for the heterogeneous and 1.72 for the homogeneous groups). In the heterogeneous groups either points were awarded nearly as equally as possible or they were almost all given to one or two members. The distribution of points by homogeneous groups tended to be more moderate, being negatively correlated, in general, to the class standing of the members on the three previous examinations. Thus the problem appears to have created greater conflict in the heterogeneous than in the homogeneous groups, but certain heterogeneous groups were able to resolve the conflict.

Interaction of sex and personality composition. Using the combined data of the present and the original studies, the interaction of sex mixture and personality composition was studied in its effects on solutions to the Mined Road and the Change of Work Procedure problems. Mixed-sex as distinct from all-male groups involve another type of heterogeneity. The results indicate that the interaction of sex and personality mixture was not the same for the two problems and suggest important differences in the requirements that the problems placed on the groups.

The results for the Mined Road problem are presented in Table 3 for groups in three categories of sex composition: All Male, One Female, and Two or Three Females. Although the means and percentages above the median for the heterogeneous groups were larger than for the homogeneous groups in all three categories of sex composition, the only significant difference occurred between the All Male groups. The differences between the homogeneous and heterogeneous groups successively diminished as the proportion of women in the group increased. (Unfortunately, only one group of all females was obtained in the 2 years' data.)

The results for the Change of Work Procedure problem, presented in Table 4, are divided only into All Male groups versus groups with One or More Females. For this problem the major difference between the homogeneous and heterogeneous groups occurred in the mixed-sex groups. Proportionately more new-

TABLE 3

HOMOGENEOUS-HETEROGENEOUS DIFFERENCES IN SOLUTION QUALITY TO THE MINED ROAD PROBLEM IN GROUPS OF DIFFERENT SEX COMPOSITION
(Combined data: 1959 and present studies)

	All Male		One Female		Two or Three Females		Total	
	Homogeneous	Heterogeneous	Homogeneous	Heterogeneous	Homogeneous[a]	Heterogeneous	Homogeneous	Heterogeneous
Percentage above median[b]	12.5*	66.7*	0.0	33.3	50.0	72.7	21.7*	52.6*
M	34.8*	75.1*	37.7	54.1	64.1	72.2	45.9*	64.3*
SD	35.8	31.3	11.3	24.4	35.3	31.9	32.1	29.3
Number of groups	8	9	7	18	8	11	23	38

[a] Data from one homogeneous group in the 1959 study that consisted of four women are included in the two or three women category.
[b] The median score was determined from the results of all four-person groups, including residual groups not assigned to either the homogeneous or heterogeneous types.
* Mean or percentage difference between homogeneous and heterogeneous groups is significant at the .05 level by *t* test or chi square test, respectively.

TABLE 4

HOMOGENEOUS-HETEROGENEOUS DIFFERENCES IN QUALITY OF SOLUTIONS TO CHANGE OF WORK PROCEDURE PROBLEM IN GROUPS OF DIFFERENT SEX COMPOSITION
(Combined data: 1959 and present studies)

Type of group	All Male Groups		One or More Females	
	Old or New Solutions	New-Modified or Integrative Solutions	Old or New Solutions	New-Modified or Integrative Solutions
Homogeneous	7	5	11*	3*
Heterogeneous	5	3	9*	17*

* This relationship is significant at the .05 level by chi square test.

modified or integrative solutions—high quality solutions—were produced by Mixed-Sex heterogeneous groups than by Mixed-Sex homogeneous groups. The proportions in both types of All-Male groups were approximately equal and were like those of the Mixed-Sex homogeneous groups. In the present study, moreover, all five of the integrative solutions were produced by Mixed-Sex heterogeneous groups (see Table 2).

The results suggest that heterogeneity of personality may be sufficient to produce high quality solutions where, as with the Mined Road, the problem presents no restraints on possible alternative solutions. The group must generate its own potential solutions to this type of problem and then decide on the best one. Where, however, as with the Change of Work Procedure, alternative choices are presented in the problem, the addition of sex mixture to the personality mixture may be required to produce high quality solutions. To arrive at a solution of high quality to this type of problem, the group must recognize that the given alternatives are inadequate and "invent" a better solution. Mixing sexes and personalities appears to have freed these groups from the restraints of the solutions given in the problem.

Acceptance of Solutions

Although on three of the four quality problems a higher proportion of heterogeneous groups produced good solutions than did the homogeneous groups, the reported member satisfactions with their solutions showed no consistent differences between the two types of groups (see Table 5). Both the mean satisfaction ratings and the percentages of groups in which all members were satisfied were closely similar on all the problems except the Mined Road problem, on which satisfaction was somewhat (but not significantly) higher for the homogeneous groups. Thus the somewhat higher quality of the heterogeneous groups' solutions was not obtained at the sacrifice of acceptance by the group members.

As might be anticipated from this result, the correlations between quality of solution and mean member satisfaction were generally rather low. These are shown for the four quantitatively scored problems in Table 6. Although all of the correlations are positive, only two of the eight achieve statistical significance.

A more important variable than quality of solutions in determining member satisfaction with the group's solution is suggested by the high correlations between the subjects' satis-

TABLE 5

SATISFACTION WITH SOLUTIONS IN HOMOGENEOUS AND HETEROGENEOUS GROUPS

	Homogeneous[a]	Heterogeneous
Mined Road Problem		
M	4.47	4.22
SD	0.37	0.43
Percentage unanimously satisfied[b]	80%	57%
Number of groups	10	21
Student Assistance Fund		
M	4.28	4.21
SD	0.67	0.67
Percentage unanimously satisfied[b]	71%	70%
Number of groups	14	20
Painters and Inspectors		
M	4.00	4.05
SD	0.70	0.57
Percentage unanimously satisfied[b]	54%	50%
Number of groups	13	16
Change of Work Procedure		
M	4.09	4.05
SD	0.81	0.44
Percent unanimously satisfied[b]	50%	35%
Number of groups	14	18
Point Distribution		
M	4.19	4.39
SD	0.69	0.41
Percentage unanimously satisfied[b]	50%	59%
Number of Groups	16	17

[a] None of the differences between group types is significant at the .05 level.

[b] To determine unanimity of satisfaction a person was considered satisfied if he checked either "very satisfied" or "quite satisfied."

faction with the amount of influence they had over the decision and their satisfaction with the solution. Table 7 shows that, with two exceptions, these correlations, both at the individual and the group levels, are all significantly positive. The correlations between the subjects' satisfaction with the solution and their actual amount of influence over the decision (as rated by the other group members) were significantly different from zero in only 3 of the 10 instances and were, in every case, smaller than the corresponding correlation between satisfaction with solution and satisfaction with influence. These correlations suggest, then, that the subjects are satisfied with the decision their group reaches if they feel that their perceived amount of influence over the decision has coincided with some self-determined level of expectation concerning the amount of influence they should have.

TABLE 6

CORRELATIONS BETWEEN QUALITY AND MEAN SATISFACTION WITH SOLUTIONS

Problem[a]	Homogeneous		Heterogeneous	
	r	Number of Groups	*r*	Number of Groups
Mined Road Problem	.01	10	.33	21
Student Assistance Fund	.49*	14	.24	20
Painters and Inspectors	.35	13	.63**	16
Point Distribution	.29	16	.26	17

[a] A Pearson coefficient could not be computed on the Change of Work Procedure since the solutions were not scored quantitatively. The fourfold contingency tables formed in each group type by combining Old and New solutions against New-Modified and Integrative and taking the number of groups above against below the median mean satisfaction showed only slight positive relationships.

* Correlation significantly different from zero at .05 level.

** Correlation significantly different from zero at .01 level.

TABLE 7

CORRELATIONS BETWEEN SATISFACTION WITH SOLUTION AND SATISFACTION WITH INFLUENCE OVER DECISION

Problem	For Individuals		For Group Means	
	Homogeneous	Heterogeneous	Homogeneous	Heterogeneous
Mined Road Problem	.54** (40)[a]	.48** (84)	.47 (10)[a]	.49* (21)
Student Assistance Fund	.63** (56)	.55** (80)	.66** (14)	.73** (20)
Painters and Inspectors	.62** (52)	.48** (64)	.74** (13)	.80** (16)
Change of Work Procedure	.46** (56)	.46** (68)	.46* (14)	.06 (17)
Point Distribution	.37** (64)	.65** (68)	.59** (16)	.88** (17)

[a] Numbers in parentheses are the number of individuals or groups on which each correlation is based.

* Correlation significantly different from zero at .05 level.

** Correlation significantly different from zero at .01 level.

DISCUSSION

The results of this study confirm, for a wider sampling of problems, the findings of the earlier one that heterogeneous groups are relatively superior to homogeneous groups in problem solving ability. Even on problems designed to produce emotional conflict, the heterogeneous groups produced solutions that were better than or at least as good as those of the homogeneous groups.

The similar personalities of members of the homogeneous groups may have provided them

with similar perspectives on each problem. Common to all of the homogeneous groups was the fact that all the members of each group had very similar self-perceptions. To the extent that similarity in self-perceptions is accompanied by similarity in the perceptions of problems and social situations, the number of different "directions" (Maier, 1930) available to each homogeneous group was much less than the number available to each heterogeneous group. With fewer directions available the likelihood was reduced that homogeneous groups would develop high quality solutions to these complex problems. In the heterogeneous groups, moreover, the conflict resulting from opposing viewpoints may have caused more complete solutions to emerge or new ones to be invented. Thus we might conclude that, although one can think of unreasonably extreme exceptions (e.g., people who do not speak the same language, adults and children), given relatively equally able people, the greater the differences in perceptions among the group members, the higher the quality of their problem solving.

Although other interpretations of these results are possible, the interactive effects of heterogeneity with respect to sex and personality on the solutions of the Mined Road and Change of Work Procedure problems seem to provide further supporting evidence for this view. Mixing sexes in the groups enhanced the quality of solutions to both problems, although in somewhat different ways for each problem.

The differing results of sex and personality mixture for the Mined Road and Change of Work Procedure problems suggest that differences in perspective are most important where the task is restrictive in its nature. But even where the problem is relatively unstructured, where the solutions must be developed by the group, the greater the differences the higher is the quality of group problem solving.

The data indicate that the higher quality of solutions associated with group heterogeneity were not obtained at the sacrifice of their acceptance by the group members. It would appear, moreover, that the very process by which the quality of group problem solving is improved, namely, the complete expression of such differences, is likely to enhance the acceptance of solutions. Having shown that a member's satisfaction with his group's solution to a problem is a function of his satisfaction with his amount of influence over the decision, we can conclude tentatively that, where differences in perspective are present among group members, the group's solution will be of highest quality and will have the most acceptance where these differences are expressed and finally utilized in making the final decision. This does not mean that each group member must have an equal amount of influence over the final decision. Rather, all divergent viewpoints must be given what the member perceives as a fair hearing by the group and be employed in making the final decision to the extent that they are seen as relevant. Not every group member has to participate actively in the discussion, as has sometimes been suggested (e.g., Coch & French, 1958). Each group member must participate, however, if he is dissatisfied with the decision as it appears to be forming and he has some alternative approach or modification to offer.

By separating the factors important for quality and acceptance of solutions in this way, one can understand how group members may be satisfied with very poor group solutions or dissatisfied with very good ones. In the first case the group may have only a few directions available among its members, but these have been expressed completely to the group's satisfaction. In the second case a variety of directions may have been available but only those which led to the correct solution were heard. Dissatisfaction with the group's decision stems from the members' feeling that their alternative viewpoints were not evaluated fairly by the group.

SUMMARY

A wider variety of problems was used to test the generality of Hoffman's (1959) earlier finding that groups composed of people heterogeneous with respect to personality were superior in solving problems to groups composed of people with homogeneous personalities. Sixteen homogeneous and 25 heterogeneous four-person groups were initially formed in the laboratory sections of an undergraduate course in the psychology of human relations. Homogeneous groups consisted of people with high positive profile correlations among their scores on the Guilford-Zimmerman Tempera-

ment Survey, heterogeneous groups of people with near-zero or negative correlations. The group members interacted weekly in case discussions, problem solving, and role playing.

Five different problems were used to test the relative abilities of the two types of groups, four with some component of quality and one only involving acceptance. The subjects were also questioned about their satisfaction with the solution and with their influence over the solution for each problem.

The following results were obtained:

1. Heterogeneous groups produced a higher proportion of high quality solutions than did homogeneous groups to three of the four problems with quality components. On the fourth problem, there was little difference and both types of groups produced poor solutions for the most part.

2. Mixed-Sex groups tended to produce higher quality solutions than did All-Male groups.

3. Satisfaction with the problem solutions was about the same in the homogeneous and heterogeneous types of groups.

4. Satisfaction with the solutions was shown to be more strongly correlated with the members' satisfaction with the amount of influence they had over the solution than with the objective quality of the solutions or with the members' actual amount of influence.

It is suggested that solutions with high quality and high acceptance can be obtained from groups in which the members have substantially different perspectives on the problem, and in which these differences are expressed and used by the group in arriving at the final decision.

REFERENCES

COCH, L., & FRENCH, J. R. P. Overcoming resistance to change. In Eleanor E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology.* (3rd ed.) New York: Holt, 1958. Pp. 233–250.

GUILFORD, J. P., & ZIMMERMAN, W. S. *The Guilford-Zimmerman Temperament Survey.* Beverly Hills: Sheridan Supply, 1949.

HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem solving. *J. abnorm. soc. Psychol.*, 1959, **58**, 27–32.

MAIER, N. R. F. Reasoning in humans: I. On direction. *J. comp. Psychol.*, 1930, **10**, 115–144.

MAIER, N. R. F. *Principles of human relations.* New York: Wiley, 1952.

MAIER, N. R. F., SOLEM, A. R., & MAIER, A. A. *Supervisory and executive development.* New York: Wiley, 1957.

(Received March 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 408–412

GROUP COHESIVENESS, COMMUNICATION LEVEL, AND CONFORMITY[1]

ALBERT J. LOTT AND BERNICE EISMAN LOTT
University of Kentucky *Kentucky State College*

By placing the concept of group cohesiveness within the framework of S-R learning theory, predictions regarding both its antecedents and consequents have been made on the basis of general behavior principles (Lott, in press). Cohesiveness is defined as *that group property which is inferred from the number and strength of mutual positive attitudes among the members of a group.* An attitude is defined, following Doob (1947), as an implicit anticipatory response having both cue and drive properties.

Basic to the present approach is the proposition that a primary condition for the development of positive attitudes toward persons is the receipt of reward in their presence. This proposition was tested in a previous study (Lott & Lott, 1960) in which confirming evidence was obtained. The present investigation is concerned, not with this determinant of cohesiveness, but with two of its consequences. It is predicted that a group's level of communication, and the tendency of its members to conform to the dominant group opinion on an issue, will be positively related to the degree of cohesiveness present within the group.

The predicted relationship between cohesiveness and communication level is derived from the assumed drive property of attitudes (Doob, 1947). Following from such an assumption is the expectation that groups composed of members with strong positive attitudes toward one another (i.e., high cohesive groups) should be characterized by a generally high drive level reflected in the amount of communication among the members. The second hypothesis follows from the expectation that the greater the cohesiveness of a group, the more can individual members function as secondary reinforcers for the behavior of one another and, thus, produce conformity.[2] This proposition is derived from the following considerations: the greater the cohesiveness of the group, the more have the members (person-stimuli) been present during past conditions of reinforcement; it is under such conditions that secondary reinforcers develop (Hull, 1951, Ch. 8); in addition, since it is assumed that the secondary reinforcing property of a stimulus depends upon the ability of the stimulus to evoke an anticipation of reward (Hull, 1952, Ch. 5), or a positive attitude, the more cohesive the group the more should its members be able to function as secondary reinforcers for one another and, thus, the more should approval of a behavior (opinion) by members serve to maintain it and disapproval to extinguish it.

In the present investigation natural, existing groups of friends were put into a semicontrolled laboratory situation. A positive correlation was predicted between measures of cohesiveness and both communication level and conformity.

METHOD

Subjects

Subjects were recruited from 15 on-going student organizations at the University of Kentucky and Kentucky State College[3] from each of which a small group of 6–10 friends was obtained. Descriptive information concerning each of the groups is presented in Table 1.

For purposes of validating a Group Cohesiveness Index, to be described below, seven groups of strangers each made up of 6–10 individuals brought together for the first time, were also used. These subjects were recruited from separate laboratory sections of introductory psychology at the University of Kentucky.

Procedure

Each group met with the investigators in a room arranged so that members sat in a circle, with one

[1] Funds for research assistance on this study were provided by the Undergraduate Participation Program of the National Science Foundation. The authors would like to acknowledge the aid, during various stages of the investigation, of Arnold Jackman, a senior at the University of Kentucky.

[2] Conformity behavior is defined as a *change in behavior (opinion) following information about the behavior of others*, where this change is in the direction of the information received.

[3] At the time of the investigation, the student body of Kentucky State College was entirely Negro. All subjects recruited from the University of Kentucky were white.

TABLE 1
DESCRIPTION OF GROUPS

Group	Type	Male: Female Ratio	*N*	School
A	Religious	4:5	9	College
B	Debating	5:1	6	College
C	Sorority	0:10	10	College
D	Academic	6:0	6	College
E	Athletic	0:7	7	College
F	Religious	3:7	10	College
G	Fraternity	9:0	9	College
H	Veteran's	10:0	10	College
I	Social	0:6	6	College
J	Nurses (1)[a]	0:6	6	University
K	Nurses (2)[a]	0:6	6	University
L	Nurses (3)[a]	0:6	6	University
M	Nurses (4)[a]	0:6	6	University
N	Social	1:5	6	University
O	Sorority	0:7	7	University
Total 15		38:72	110	

[a] The four groups of nurses were separate friendship groups from within the larger freshman nursing class.

investigator (E_1) slightly behind the circle and the other (E_2) at a table apart from the group. Each subject was given an identifying letter to pin on his (her) shirt. The letters were distributed in alphabetical order, clockwise around the circle, starting from the subject seated to the left of E_1.

The purpose of the investigation was explained to the subjects as being that of obtaining the opinions of students on a subject of current interest to educators. Before the discussion topic was introduced, however, the group members were told: "We would like to know how each one of you feels about every other member of your group present here now," and the following instrument was administered: a booklet containing rating forms on which each individual could rate every other on a nine-point scale, ranging from "dislike very much" (1) to "like very much" (9). The booklet contained a separate page for each member, other than the rater, whose identifying letter simply appeared on the booklet's cover page. The last page of the booklet called for the rater's "*attitude toward the group as a whole.*" The person rated on each page was identified by letter and, since no record was made of the names associated with letters, subjects were assured of complete anonymity. The subjects were also informed of the fact that the sequence of letters within the booklets had been randomized so that each individual would be rating different members of the group at different times.

After the booklets were collected, E_1 presented the topic for discussion as follows:

It is being said, by many professional educators as well as other general observers, that today's average college student is interested only in himself and takes little or no interest in the problems facing his community, his country or his world. It is being said that the college student is too contented, is lazy, and self-centered. We would like to have your opinions on this question. Do you agree; do you disagree; why? Because we want only your opinions, we will *not take part* in the discussion. We'll stop you at the end of half an hour.

While E_1 took content notes, E_2 recorded communication frequency by tallying separate remarks made within 5-second periods on a standard form suggested by Gibb, Platts, and Miller (1951). A separate tally was recorded for each comment made within a 5-second interval, regardless of its total duration. Thus, if one subject spoke continuously for 5 seconds, only one tally was made; if several subjects spoke during the same 5-second interval, one tally was made for each speaker; and if one subject spoke for 15 seconds, three tallies were made, one for each 5-second interval utilized by him. Each tally identified the speaker by letter.

At the end of 30 minutes, the discussion was terminated and an opinion scale was administered to each subject preceded by the following instructions: " . . . Sometimes views expressed openly in discussion are not quite the same as when expressed privately. We would therefore like, now, to get an expression of opinion from each one of you individually. . . . These opinions are anonymous. Please do not put either your name or your letter on the paper, just your opinion. Let me repeat the statement: . . . ." The opinion scale consisted of nine points from "strongly disagree" (1) to "strongly agree" (9). The scales were collected in such a way as to preserve the identifying letter order of the seated subjects, thus, making it possible to later match the responses to this scale with responses made to one administered subsequently.

Following collection of the scales, subjects were requested to maintain a brief period of complete silence and thoughtfulness. The atmosphere of this period was augmented by a recorded musical background provided by the first 5 minutes of Prokofiev's *Seventh Symphony*. During this period E_2 tabulated the opinion scale responses which were reported to the group at the end of 5 minutes. The reported results were always *directly opposite* to the actual findings. Thus, a group which had actually indicated a tendency towards disagreement with the statement was told that the results showed agreement, and vice versa.

After the "results" were presented to the group, the musical interlude was explained as follows: "There is evidence to suggest that sometimes just sitting quietly permits a person to think of things which he did not think of earlier, in the middle of a discussion. This might have happened to some of you. Just in case it did, we would like to give you another chance to express your opinion on the issue which the group discussed. Your opinions are, again, completely anonymous . . ." A second opinion scale was then administered and collected in the same manner as described above. The subjects were asked to make any comments they wished on the back of their opinion blanks.

Measures

From the data gathered by means of the instruments described above, measures of cohesiveness, communication level, and conformity were obtained for each group.

Cohesiveness. From the booklets which were administered at the beginning of the experimental session, the attitude of each group member toward every other member was ascertained. In obtaining an index of group cohesiveness from these individual responses, each member was paired with every other member and their ratings of each other were averaged. To simplify this procedure, the original nine-point scale was converted to one ranging from −4 (Point 1) through 0 (Point 5) to +4 (Point 9). Thus, if Person A had circled Point 8 on the scale for Person B and Person B had circled Point 5 for Person A, their "mutual score" or pair average was +1.5. The Mean of all the "mutual scores" or pair averages was then obtained as the Group Cohesiveness Index (GCI).

This measure of cohesiveness follows directly from the conceptual definition presented earlier. It takes into account the strength and direction of attitudes of *each member toward every other* and, thus, reflects the mutual attitudes among the group members.

Communication level. A description of the tally system used in recording frequency of communication within each group was given above. From the tallies, a group measure of communication level was obtained by calculating the Median number of verbalizations made by the individual group members. A verbalization is defined as a single remark, regardless of its duration. A Median was utilized as the group score because it is not influenced by extreme variations in individual group member scores which probably reflect differential "verbalness" of individual participants.

Conformity. The two sets of opinion scales on the discussion topic, which were obtained from each group, were compared. Since the second scale followed the presentation of contrived information regarding the general direction of results on the first scale, a post-information opinion could be compared with a pre-information opinion for every group member. The percentage of group members who changed their opinion in the direction of the contrived information was taken as the measure of conformity within each group.

RESULTS

Validity of the GCI

The validity of the GCI was tested by comparing the scores obtained on the 15 natural friendship groups with those obtained on 7 groups of strangers. White's rank test (Edwards, 1954) was used to test the significance of the difference between groups; the obtained T' value of .31 is significant at the .01 level (two-tailed). The group scores on the GCI for the strangers ranged from .8–1.7, in contrast to those for the natural groups which ranged from 1.5–3.6. The ability of the GCI to clearly differentiate between groups of strangers and groups of friends, where differences in cohesiveness are to be expected, was, thus, established.

An additional validation test of the GCI was made possible by examining responses to the last question which had been asked of the subjects in the attitude-toward-group-members booklet: Will you now check your attitude towards the *group as a whole*? (on a nine-point scale). Mean group scores to this question, which ranged from 7.3–9.0, were correlated with CGI scores, resulting in a tau of .74, significant at less than the .001 level.

TABLE 2

SCORES OBTAINED BY GROUPS ON MEASURES OF COHESIVENESS, COMMUNICATION, AND CONFORMITY

Group	GCI	Group Scores	
		Communication Level	Proportion of Conformity Responses
O	3.6	44.0	.29
J	3.5	88.5	.50
I	3.2	16.5	.33
K	3.1	42.0	.50
G	3.1	6.0	.22
C	3.1	5.5	.10
N	2.9	32.0	.17
M	2.8	18.5	.17
B	2.7	15.0	.17
D	2.3	15.0	.33
F	2.2	14.5	.30
L	2.1	52.0	.00
H	2.1	11.5	.00
A	1.8	6.0	.11
E	1.5	9.0	.00

Cohesiveness and Communication Level

The scores obtained by each natural friendship group on these two variables are presented in Table 2. The relationship between cohesiveness and communication level was tested by means of Kendall's rank correlation coefficient, tau (Siegel, 1956). The obtained correlation of .43 is significant at the .02 level (two-tailed).

Cohesiveness and Conformity

The conformity score for each group is also given in Table 2. A tau analysis of the relationship between conformity and group cohesiveness yielded a coefficient of .54, significant at less than the .01 level of confidence (two-tailed).

DISCUSSION

The results of this investigation indicate that significant positive relationships exist between group cohesiveness and both communi-

cation level and conformity behavior. Therefore, given knowledge of the degree of a group's cohesiveness, as the concept has been here defined and measured, reliable predictions of two significant types of group behavior can be made.

The authors are not aware of other reports of a direct relationship between the quantitative level of communication within a group and its degree of cohesiveness, although a qualitative relationship between high and low cohesiveness and communication patterns has been shown (Back, 1951). Both Newcomb (1953) and Festinger (1950), however, have suggested that frequency of communication between persons should vary directly with the degree to which they are attracted to one another (Newcomb) or to the group (Festinger). The present approach considers communication as only one manifestation of a group's general drive level, to which the drive components of positive attitudes among group members contribute. Other reflections of drive level, such as task persistence, productivity, etc., are also expected to show positive relationships with cohesiveness.

Empirical support for a positive relationship between group cohesiveness and conformity among group members can be found in the literature (e.g., Berkowitz, 1954; Festinger, Schachter, & Back, 1950; Schachter, 1951), but negative results have also been reported (Bovard, 1953; Downing, 1958). Although the matter cannot be gone into fully here, the present approach would not predict a positive relationship between cohesiveness and conformity in all instances, but only in those situations where the stimulus features are similar, to some degree, to those which were present when group members acquired their secondary reinforcing characteristics. The principle of stimulus generalization, in other words, should be employed to predict the effectiveness with which group members can act as secondary reinforcers for one another in diverse situations.

Evidence supporting the validity of the GCI used in this investigation has already been presented. In view of the high correlation reported between the GCI and responses to the question regarding attitude toward *group as a whole*, the question arises as to whether the latter might not do just as well as a measure of cohesiveness, as the former. Just such a simple measure has been suggested by Gross and Martin (1952). Points that might be considered in answering this question are the following: examination of the rating that individuals made of their groups-as-a-whole shows a marked tendency for use of the upper portions of the scale whereas a similar restrictive tendency is not found in ratings made of individual group members; the Gross-Martin Question predicted more poorly to the dependent variables of the present investigation;[4] and the GCI is directly derived from our conceptual definition.

SUMMARY

The predictions that both communication level and conformity behavior within groups vary positively with the degree of group cohesiveness were clearly confirmed. Measures of the three variables under investigation were obtained, in a laboratory setting, on 15 natural friendship groups, each of which engaged in a 30-minute discussion of a presented topic. Data pertaining to the validity of a cohesiveness index, derived from the conceptual definition of cohesiveness, were also obtained.

Since the hypotheses tested and supported in this investigation are part of a formulation of group cohesiveness within the framework of learning theory, the obtained results lend support to attempts to predict group behavior on the basis of more general behavior principles.

REFERENCES

BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–24.

BERKOWITZ, L. Group standard, cohesiveness, and productivity. *Hum. Relat.*, 1954, **7**, 509–519.

BOVARD, E. W., JR. Conformity to social norms and attraction to the group. *Science*, 1953, **118**, 598–599.

DOOB, L. W. The behavior of attitudes. *Psychol. Rev.*, 1947, **54**, 135–156.

DOWNING, J. Cohesiveness, perception, and values. *Hum. Relat.*, 1958, **11**, 157–166.

EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

[4] A tau of .31 was obtained for communication and a tau of .34 for conformity. Neither is significant at the .05 level.

FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.

FESTINGER, L., SCHACHTER, S., & BACK, K. *Social pressures in informal groups*. New York: Harper, 1950.

GIBB, J. R., PLATTS, GRACE, & MILLER, LORRAINE. *The dynamics of participative groups*. St. Louis: John Swift, 1951.

GROSS, N., & MARTIN, W. E. On group cohesiveness. *Amer. J. Sociol.*, 1952, **57**, 546–554.

HULL, C. L. *Essentials of behavior*. New Haven: Yale Univer. Press, 1951.

HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.

LOTT, BERNICE E. Group cohesiveness: A learning phenomenon. *J. soc. Psychol.*, in press.

LOTT, BERNICE E., & LOTT, A. J. The formation of positive attitudes toward group members. *J. abnorm. soc. Psychol.*, 1960, **61**, 297–300.

NEWCOMB, T. An approach to the study of communicative acts. *Psychol. Rev.*, 1953, **60**, 393–404.

SCHACHTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190–207.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

(Received March 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 413–421

SOCIAL DEPRIVATION AND RIGIDITY IN THE PERFORMANCE OF FEEBLEMINDED CHILDREN[1]

EDWARD ZIGLER
Yale University

THE view that feebleminded individuals are more rigid than normal individuals of the same MA has been advanced by Lewin (1936) and Kounin (1939, 1941a, 1941b, 1948). Recent findings (Plenderleith, 1956; Stevenson & Zigler, 1957), however, have failed to support the Lewin-Kounin formulation. In evaluating the disagreement of these findings with those of Lewin and Kounin, Stevenson and Zigler (1957) noted that in the later studies the subject was required to learn a response, while in the earlier studies, the subject was required to perform a response in compliance with instructions.

The hypothesis was advanced that the differences in performance of normal and feebleminded subjects reported by Lewin and Kounin were related to motivational differences between the groups, rather than to differences in cognitive rigidity. This hypothesis was based on the assumption that institutionalized feebleminded children tend to have been relatively deprived of adult contact and approval and hence have a higher motivation to secure such contact and approval than do normal children. This assumption appears congruent with the view advanced by other investigators that both institutionalized feebleminded and institutionalized normal individuals exhibit an increased desire to interact with adult figures (Clark, 1933; Sarason, 1953; Skeels, Updegraff, Wellman, & Williams, 1938).

Zigler, Hodgden, and Stevenson (1958) employed the above-stated assumption in deriving the hypothesis that interaction with an adult and adult approval provide greater reinforcement for the responses of institutionalized feebleminded subjects than they do for those of normal subjects. Predictions concerning the performance of feebleminded and normal subjects on instruction-initiated tasks made on the basis of this hypothesis were supported. The findings of Gewirtz and Baer (1958a, 1958b) provide further evidence for the relationship between social deprivation and the effectiveness of adult approval as a reinforcer.

The studies of Zigler et al. (1958) and Gewirtz and Baer (1958a, 1958b) lend support to the view that the reported differences in the incidence of "rigid" behavior between normal and feebleminded subjects of the same MA are a function of the greater social deprivation experienced by feebleminded subjects rather than of the inherent rigidity of the feebleminded. A direct test of this position is suggested. If the rigid behaviors found in feebleminded subjects are a result of the social deprivation the subjects have experienced, then within a feebleminded population a relationship should exist between the degree of deprivation experienced and the amount of rigidity manifested. The following hypothesis is proposed:

> The greater the amount of pre-institutional social deprivation experienced by the feebleminded child, the greater will be his motivation to interact with an adult, making such interaction and any adult approval or support that accompanies it more reinforcing for his responses than for the responses of a feebleminded child who has experienced a lesser amount of social deprivation.

The methodological difficulties entailed in an adequate test of this hypothesis have been discussed at length elsewhere (Zigler, 1958). The problem of adequately defining "social deprivation" merits special attention. As others have noted (Gewirtz & Baer, 1958a), the concept of social deprivation has been applied loosely to certain events in early child-

[1] This report is adapted from a dissertation submitted to the graduate faculty of the University of Texas in partial fulfillment of the requirements for the degree of doctor of philosophy. Harold Stevenson initiated with the author a series of experiments which led directly to this study. To Louis Moran, who directed the study, and to Harold Stevenson, the writer is greatly indebted for their continuous constructive criticism.

The study was made possible through the cooperation of Malcolm J. Farrell, Superintendent, and the staff of the Walter E. Fernald State School at Waverly, Massachusetts. Special thanks go to Catherine E. Chipman, principal psychologist, and to John H. Davidson, staff psychologist, for their conscientious work in rating the subjects on the social deprivation scale used in the study.

hood which in turn are characterized as being antecedent to certain social behaviors (Bowlby, 1951; Goldfarb, 1945; Ribble, 1944; Spitz, 1945). Gewirtz and Baer (1958a, 1959b) avoided such looseness by equating "social deprivation" to a brief period of social isolation. They thus reduced deprivation to a short-term operation often employed in general behavior theory in which deprivation means "the unavailability to an organism of a given reinforcer for a period sufficient to effect an increase in behaviors for it." As Gewirtz (1957) has pointed out in another paper, however, such a short-term contemporaneous model is not the appropriate one when dealing with the concept of deprivation as employed by such investigators as Spitz, Ribble, Goldfarb, and Levy. To these latter investigators deprivation refers to long-time deficits and hence an appropriate definition of deprivation must reflect conditions operative through wide segments of time.

In dealing with the social deprivation that feebleminded children have experienced, a measure that would reflect such long-term deficits is required. Since the population of events which constitutes social deprivation has never been adequately delimited, the procedure employed in the present study was that of allowing raters to evaluate the entire preinstitutionalization history of the individual for degree of social deprivation, while at the same time having them note the specific factors in the history which determined the rating. It was hoped that this procedure would result in not only an assessment of social deprivation for each subject, but also in the extraction of a universe of specific events, the occurrence or nonoccurrence of which could be employed as a definition of social deprivation in future investigations.

The present study used a two-part game very similar to one employed by Zigler et al. (1958). On Part I, the subject was instructed to make one type of response and was allowed to continue making this response until satiated or until he had made the maximum number of responses allowed by the game. On Part II, the subject was instructed to make a different type of response to the same stimuli and again was allowed to perform until satiated or until he had made the maximum number of responses allowed by the game. Two conditions of reinforcement were employed. In one condition the subject performed Part I of the game under Nonsupport (attention only) and then performed Part II under Support (supportive comments). This was designated the Nonsupport to Support condition (NS–S). In a second condition the subject performed Part I of the game under Support and Part II under Nonsupport. This was designated the Support to Nonsupport condition (S–NS).

From the hypothesis advanced in this study, and by employing the above game and conditions of reinforcement, certain predictions can be made. These predictions and their rationales are presented below:

1. The more socially deprived subjects perform longer on the game than the less socially deprived subjects under both NS–S and S–NS conditions. This prediction was derived directly from the hypothesis that the greater the amount of social deprivation experienced, the greater is the motivation to interact with adults. Since both NS (attention) and S (supportive comments) are reinforcing, they both should have a greater reinforcing effect on the more socially deprived subjects than on the less socially deprived subjects.

2. The more socially deprived subjects make a lower proportion of errors than the less socially deprived subjects. If, as is hypothesized, more socially deprived subjects have a higher motivation to secure adult interaction and approval, they should show a greater compliance with the experimenter's instructions. This greater compliance should lead to fewer errors.

3. The more socially deprived subjects more frequently make the maximum number of responses on the game. This prediction was derived from two sources: since it was hypothesized that the social reinforcers being offered would be more reinforcing for the responses of these subjects who have experienced the greater amount of social deprivation, the more likely it should be that these social reinforcers would continue to be reinforcing throughout the maximum number of responses allowed; and, if, as was assumed, higher motivation to secure adult interaction and approval results in greater compliance with the experimenter's instructions, the more socially deprived subjects might feel that such compliance involved completing the game.

4. The increase in time spent in Part II over that spent in Part I of the game is greater for the more socially deprived than for the less socially deprived subjects under both NS–S and S–NS conditions. This prediction for the NS–S condition is supported by findings of an earlier study (Zigler et al., 1958), which indicated that the S condition is more reinforcing than the NS condition. It would follow that this increase in reinforcement would have a greater effect on the more socially deprived subjects due to their greater motivation to attain this reinforcement. The prediction for the S–NS condition was based on the assumption that a child switched from an S to an NS condition would evidence an increase in output under the NS condition

in the hope that such an increase would lead to a reinstatement of S. Since the more socially deprived subjects were considered to have the higher motivation for the reinstatement of S, it was expected that they would show a greater increase in time spent on Part II than would the less socially deprived subjects under the S–NS condition

In addition to testing these predictions, the present study was directed toward an examination of the differences in performance related to the two conditions of reinforcement. While it was assumed that there would be an increase in the number of responses in Part II over that in Part I under both the NS–S and S–NS conditions, other differences in performance that emerged under the two conditions also were of interest. An examination of such differences was undertaken in an effort to furnish a conceptual framework within which to interpret the finding of Zigler et al. (1958) of an increase from Part I to Part II when both parts were given under an identical S condition.

METHOD

Subjects

The subjects were 60 feebleminded children residing at the Walter E. Fernald State School in Waverly, Massachusetts. The conditions for inclusion in the sample were that the child have a diagnosis of mental deficiency, familial type; be free of any gross motor or sensory disturbances; have a CA between 8.0 and 14.0 years, have an MA between 4.0 and 8.0 years; and have been a resident of the school for at least one month. The school's admission ledger was employed to select the sample. Beginning with the most recent admissions, every individual who met the above requirements was included in the sample. Names were chosen in order of admission with the sampling procedure continuing until a group of 30 boys and 30 girls had been formed. The birth date, MA, and IQ for each child were taken directly from the school's records. Each child's IQ was used to adjust his MA for the time interval that had intervened between his MA rating and his participation in the present study. In all cases the 1937 revision of the Stanford-Binet had been used to obtain the original MA ratings.

Experimental Game (Marble-in-the-Hole)

The game was repetitive and monotonous and was constructed to reduce relatively quickly the subject's desire merely to "play the game," thus allowing the assumed difference in motivation for interaction with the experimenter to become the dominant variable in determining performance. The game was extremely simple, and successful performance depended primarily upon compliance with the experimenter's instructions rather than upon learning. The rules of the game allowed the subject to stop playing whenever he wised. Although the game had no defined goal, there was the implicit but distant goal of inserting all 400 marbles and thus emptying a container. This game was very similar to the Peg-in-the-Hole game used by Zigler and his colleagues (1958).

In Marble-in-the-Hole, the subject's task was to insert a marble of one color into one hole and a marble of another color into a second hole of an enclosed box. The apparatus included: a green box 9 × 6 × 6 inches with two small square holes along the middle of the lid of the box and separated by 2 inches; 400 marbles, 200 blue and 200 orange; a red open container 3 × 4 × 4 inches to hold the marbles.

Part I. The container filled with 400 marbles and the box were placed in front of the subject. The experimenter said:

> This is a game that we call "Marble-in-the-Hole." I'll tell you how to play it. You see these marbles. Some of them are blue and some of them are orange. They go in these holes. The blue ones go in the front hole and the orange ones go in the back hole. (The experimenter points to the appropriate holes.) Now show me a blue marble. Put it in the hole it goes in. Now show me an orange marble. Put it in the hole it goes in. You can put as many marbles in the holes as you want to. You tell me when you want to stop and we'll play a new game.

The subject then played the game until he indicated that he wished to stop either by telling the experimenter that he wanted to stop; or not inserting a marble for 30 seconds.

Part II. After the subject indicated that he wished to stop, the experimenter emptied the marbles from the box back into the container and then said:

> Now I'll tell you how to play *this* game. This time we put the orange marbles in the front hole and the blue marbles in the back hole. Put an orange marble where it goes. Now put a blue marble where it goes. You can put as many marbles in the hole as you want to. You tell me when you want to stop.

The subject was allowed to continue until he told the experimenter that he wished to stop or he did not insert a marble for 30 seconds.

The color of the marbles and the particular holes into which they were to be placed were randomized among the subjects so that all possible combinations and switches were used. The maximum number of marbles that could be inserted on each part of the game was 400.

Procedure

Performance variable. The subjects were tested individually. The experimenter brought the subject into a room which contained a small table and two chairs and told him:

> We're going to play some games today. They are fun games and I think that you will like them. You will be able to play each game as long as you want to.

The subject was seated at the table and the experimenter arranged the materials and gave the instructions for Part I of the game.

Each subject played one part of the game under an NS condition and the other part under an S condition. In the NS condition the instructions were given and

the experimenter did not interact further with the subject during the games. No verbal support was given to the subject and care was taken to avoid nodding, smiling, or giving other types of nonverbal support. In the S condition the instructions were given and the experimenter continued after this to interact with the subject. This interaction included such nonverbal support as smiles and nods and verbal support in the form of comments about the subject's performance. The following comments were used: "You really know how to play this game," "That's fine," "Very good," "You really know how to play Marble-in-the-Hole, don't you?" These statements were made in a random order, one after every tenth marble was inserted.

Half the subjects played Part I of the game under the NS condition and played Part II under the S condition (NS–S group). The other half of the subjects played Part I of the game under the S condition and played Part II under the NS condition (S–NS group). The subjects were systematically assigned to the two groups to control for MA, CA, length of institutionalization, and sex in each group. The composition of the two groups may be seen in Table 1.

The experimenter recorded the number of responses, the total time spent, and the number of errors for each part of each game. At the conclusion of the experiment, every child was told he had done very well.

Social deprivation variable. Two experienced psychologists (the school's principal psychologist, with 29 years of experience at the school, and a staff psychologist, with 3 years experience at the school) were asked to read the social history of each subject, available in the school's files, and then independently to rate each subject on a social deprivation scale. Each rating was made on a 60-cm. continuous line evenly subdivided for reference into six areas, "very protected," "protected," "slightly protected," "slightly deprived," "deprived," and "very deprived." The social deprivation score for each subject was the average of the ratings of the two judges. The possible range of scores was from 1 (extreme of "very protected") to 60 (extreme of "very deprived").

The judges were not instructed as to what consituted "deprivation" or "protection," beyond being told that it was concerned with the amount and quality of the interaction that the child had with adult figures. The judges were asked to list for each subject the specific factors in the subject's social history which influenced their rating for that subject. This last procedure was included in the hope that such a listing would uncover specific factors which eventually could be employed in purifying and making more operational the social deprivation variable.

The judges were cautioned against using any data other than that contained in the pre-institutionalization history of the child in making their ratings. They had no knowledge of the subject's performance on the experimental task, and the experimenter had no knowledge of the subject's social deprivation ratings until all subjects had completed the experiment.

TABLE 1

COMPOSITION OF THE NS–S AND S–NS GROUPS OF SUBJECTS

Group	*N*	Mean CA (in yr.)	Mean MA (in yr.)	Mean IQ	Mean Number Months Institutionalized	Mean CA (in yr.) Upon Admission
Nonsupport–Support	30	10.9	6.3	59.0	23.4	8.9
Support–Nonsupport	30	10.7	6.1	57.3	22.9	8.8

RESULTS

The game used in the study proved to be of interest to the subjects. Although there was great variability in the length of time the subjects played the game, there were no noticeable differences among the subjects in their initial attitudes toward it. No subject mentioned that Part II of the game was similar to Part I, and all responded to Part II as if it were a new game.

Social Deprivation Variable

The correlation between social deprivation ratings of the two judges was .77. The judges reported that it was difficult to rate certain subjects because of the meager amount of information contained in their social histories. Both judges commented upon the high proportion of subjects who had experienced considerable amounts of social deprivation. The combined ratings of the judges placed 18 subjects in the "very deprived" range of the social deprivation scale, 19 subjects in the "deprived" range, 10 subjects in the "slightly deprived" range, 1 subject in the "protected" range and 0 in the "very protected" range. Judges were instructed also to list for each subject the specific factors in the subject's social history which influenced their rating for that subject. The most frequent factors listed, and the frequency with which one or both judges noted them are presented in Table 2. The mean of the social deprivation scores of the NS–S group did not differ significantly from that of the S–NS group.

Since the purpose of the study was to compare the performance of the more socially deprived subjects with that of the less socially deprived subjects, subjects were divided at the median of the social deprivation scores (44.5). Two groups were formed in this manner, a high socially deprived group (HD group) and a low socially deprived group (LD group). The range of social deprivation scores

TABLE 2

FREQUENCY WITH WHICH ONE OR BOTH JUDGES NOTED A SPECIFIC FACTOR AS INFLUENCING THEIR RATINGS

Factors in Social History	*N*
Child comes from an orphanage or has lived in several homes	20
Child's parents are divorced or separated	16
Child was abused (physical punishment, sex play) or neglected (inadequately fed or clothed) to the extent that legal action was taken to remove him from his home	14
Child comes from original home in which he experienced considerable abuse or neglect, but no legal action was taken to remove child from the home	13
Mother or father institutionalized (mental hospital, colony for the feebleminded, or jail)	12
Child is illegitimate	6
Mother verbalized her negative feelings toward the child	3

was from 18 through 44 for the LD group and from 45 through 56 for the HD group.

Analysis of Predictions

The first prediction was that the more socially deprived subjects would spend a greater amount of time on the game than would the less socially deprived subjects. The difference in the average time spent on the game by the HD group (1,039.8 seconds) and LD group (655.5 second) was 384.2 seconds, which is a significant difference ($t = 2.39$, $p < .01$). (In all analyses of predictions one-tailed tests of significance were employed.)

The findings relative to the second prediction, that the more socially deprived subjects would make a lower proportion of errors than the less socially deprived subjects, were not found to be significant. Error scores were computed by dividing the total number of errors each subject made by the total number of marbles he inserted. Because of the markedly skewed distribution of error scores, chi square was employed to test this prediction. The median of the error scores (.015) was used as the dividing point in setting up a 2 × 2 contingency table. In the LD group, 12 subjects had an error score at or below the median, while 19 subjects had an error score above the median. In the HD group, 18 subjects had an error score at or below the median, while 11 subjects had an error score above the median. Although this difference was in the predicted direction, it did not reach the .05 level of significance ($X_c^2 = 2.40$; $p < .10$).

The third prediction was that the more socially deprived subjects would more frequently make the maximum number of responses on the game. In the LD group, 8 of the 31 subjects made the maximum number of responses the game allowed (inserting 400 marbles) on either Part I or Part II of the game, while 16 of the 29 subjects in the HD group did so. This difference was significant ($X_c^2 = 4.23$; $p < .025$) and confirmed the third prediction.

The fourth prediction, that the increase in time spent on Part II over that spent on Part I of the game would be greater for the more socially deprived subjects than for the less socially deprived subjects, also received confirmation. The subjects who made the maximum number of responses on both Part I and Part II (five in the LD group and five in the HD group) were not included in the test of this prediction since their performance precluded any meaningful increases or decreases from Part I to Part II. The difference in the average time increase from Part I to Part II by the HD group (240.3 seconds) and the LD group (113.7 seconds) was 126.6 seconds, which was significant ($t = 1.87$; $p < .05$). The performance on the game by the HD and LD groups is shown in Figure 1, in which the average amount of time spent on Part I and Part II is presented.

In a further analysis of the fourth prediction, differences between groups in the number of subjects who did not complete Part I, but who did complete Part II, were noted. In the

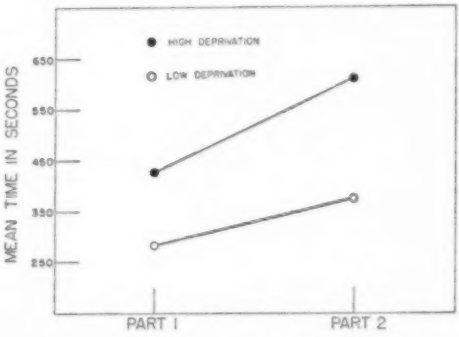

FIG. 1. Mean time in seconds spent on Part 1 and Part 2 of the game by the HD and LD groups.

LD group, 23 subjects did not perform the maximum number of responses on either part, while 3 subjects made the maximum number of responses on Part II after making less than the maximum number of responses on Part I. In the HD group, 13 subjects did not perform the maximum number of responses on either part, while 11 subjects made the maximum number of responses on Part II after making less than the maximum number of responses on Part I. This difference in the number of subjects in the HD and LD groups who did not complete Part I but who did complete Part II is significant beyond the .01 level ($X_c^2 = 5.68$).

Thus three of the four predictions were confirmed. The more socially deprived subjects were found to spend a greater amount of time on the game, to more frequently make the maximum number of responses allowed by the game, and to evidence a greater increase in time spent on Part II over that spent on Part I of the game. The prediction that the more socially deprived subjects would make a lower proportion of errors was not confirmed, though the findings were in the predicted direction.

Other Findings

As noted earlier, the study also was directed toward an examination of the relationship of the performance measures to the two conditions of reinforcement employed. Performance under the two conditions of reinforcement is shown in Figure 2, in which the average amount of time spent on Parts I and II of the game is presented. As can be seen in Figure 2, subjects spent more time on Part II than on Part I under both conditions of reinforcement. The mean difference between time spent on Part I and Part II is significant for the NS–S group (161.1 seconds) beyond the .01 level ($t = 4.35$), while the mean difference for the S–NS group (114.2 seconds) is significant beyond the .05 level ($t = 2.25$). The difference (46.9 seconds) between these differences is not significant ($t < 1.00$), indicating that the magnitude of increase from Part I to Part II did not differ under the NS–S and the S–NS conditions.

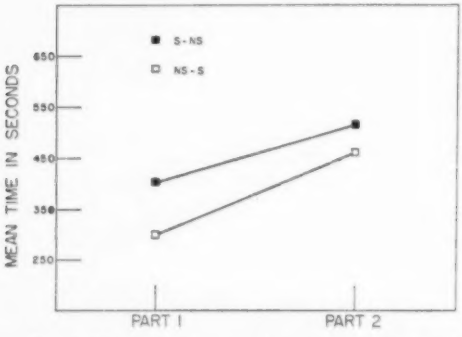

FIG. 2. Mean time in seconds spent on Part 1 and Part 2 of the game by the S–NS and NS–S groups.

The difference in the average time spent on Part I between the NS–S (300.3 seconds) and the S–NS (404.2 seconds) groups was not significant ($t = 1.29$; $p > .20$). The difference in the average time spent on Part II between the NS–S (461.0 seconds) and the S–NS (518.1 seconds) groups also was not significant ($t < 1.00$). The difference in the average total time (Part I and Part II) spent on the game between the NS–S (761.3 seconds) and the S–NS (922.3 seconds) groups was not significant ($t < 1.00$). Therefore, it is concluded that there was no evidence to indicate the groups tested with a sequence of NS–S and S–NS differed significantly on time spent on Part I, time spent on Part II, increase in time spent on Part II compared to Part I, nor on the total time spent on the game.

Relation of Social Deprivation to Other Variables

The characteristics of the HD and LD groups were examined further to ascertain whether the two groups differed on variables other than the performance measures employed in the predictions. The composition of the LD and HD groups is presented in Table 3. An analysis was made of the differences between the two groups in regard to CA, MA, IQ, length of institutionalization, and age upon institutionalization. The mean IQ of the

TABLE 3
COMPOSITION OF THE HIGH AND LOW SOCIALLY DEPRIVED GROUPS

Group	*N*	Mean CA (in yr.)	Mean MA (in yr.)	Mean IQ	Mean Number Months Institutionalized	Mean CA (in yr.) Upon Admission
High	29	10.3	6.3	61.0	22.9	8.4
Low	31	10.9	6.1	56.0	23.4	9.2

subjects in the HD group was significantly higher than that of the LD group ($t = 2.10$; $p < .05$). The t's associated with the differences in the other factors were all less than 1.00, with the exception of that associated with age upon institutionalization, which was 1.55 ($p > .10$). In light of the higher mean IQ of the HD group, a correlation was run between the subject's total time score and his IQ. This correlation was not significant ($r = -.09$; $p > .20$), indicating that while a relationship exists between the time scores and the measure of social deprivation, there is no evidence of a relationship between time scores and IQ.

DISCUSSION

The hypothesis advanced in this study was that interaction with an adult and adult approval provide greater reinforcement for the responses of those feebleminded subjects who have experienced a greater amount of social deprivation than for those who have experienced a lesser amount of social deprivation. The results lend support to the hypothesis in that three of the four predictions derived from the hypothesis were confirmed. It may be concluded that these findings offer further support for the view advanced in earlier studies (Stevenson & Zigler, 1957; Zigler et al., 1958) that the rigid behavior observed in feebleminded subjects is probably a product of higher motivation to maintain interaction with an adult and to secure approval from him through compliance and persistence. The results of the present study also offer evidence that the feebleminded subjects' higher motivation to interact with an adult is related to the greater social deprivation such subjects have experienced. Furthermore, individual differences among the feebleminded in the frequency with which they exhibit persistive and/or compliant behavior also can be related to differences in the amount of social deprivation experienced. Since the persistence and compliance exhibited by feebleminded subjects have been found to be related to social deprivation, the prediction can be made that such persistence and compliance would also be shown by subjects of normal intelligence who have also experienced social deprivation.

The experimental method may have obscured differences between the high and low socially deprived groups in their tendency to make errors. The subjects were not informed of their errors and received the reinforcement inherent in the completion of even incorrect responses and continued to receive the experimenter's attention or support following errors. Another factor that may have obscured differences between the two groups in their tendency to make errors was the significantly longer performance of the HD subjects, making the performance of these subjects more susceptible to fatigue and transfer effects.

Discussion of Other Findings

Another finding of interest is the similarity in the performance of subjects under both of the conditions of reinforcement employed. Both the subjects who played Part I of the game under the NS condition and Part II under the S condition and the subjects who played Part I under the S condition and Part II under the NS condition showed a significant increase in time spent on Part II over that spent on Part I of the game. This gives strength to the assumption that there would be an increase in time spent on Part II over Part I under the NS–S condition due to the greater reinforcing effect of support. It also strengthens the assumption that there would be an increase in time spent on Part II over Part I under the S–NS condition due to the belief of the subjects that continued performance on Part II may result in a reinstatement of support. However, the two conditions of reinforcement produced no significant differences in the increase of time spent on Part II of the game over that spent on Part I, on the time spent on Part I alone, on the time spent on Part II alone, or on the total time spent on the game.

The assumptions underlying the expected increase from Part I to Part II of the game in the present study do not appear capable of explaining the finding by Zigler et al. (1958) that feebleminded subjects show an increase in Part II over Part I of the game when both parts are played under identical conditions of support. It appears that another process related to the increase from Part I to Part II of the game may be operating. The assumption can be made that institutionalized feebleminded subjects begin the game with a greater amount of ambivalence toward it and toward

the experimenter than do normal subjects. Sarason and Gladwin (1958) have discussed the possibility that one of the major effects of institutionalization is an "avoidance and fear of new problem-solving situations" (p. 243). This suggests that institutionalized feebleminded subjects would be characterized by certain negative feelings toward the experimental situation.

Assuming that institutionalized feebleminded subjects have such negative feelings, one may advance the hypothesis that such subjects begin playing Part I of the game with a relatively high positive tendency to respond due to a greater motivation to interact with and to gain the approval of an adult which, in turn, stems from the greater social deprivation they have experienced. At the same time, they begin playing Part I of the game with a negative tendency to respond due to a wariness or suspiciousness of adults which stems from the negative experiences, perhaps of a painful nature, which they have had at the hands of adults. If Part I is given under an S condition, it is likely that the negative tendencies would decrease. The deprived child learns during Part I that the experimenter is not like other strange adults he encounters in the institution who often initiate painful experiences (physical examinations, shots, etc.) with supportive comments. Then when the deprived child is switched to Part II of the game, whether Part II is given under S or NS, he meets it with a positive tendency which has been reduced much less than his negative tendency. The result would be that he performs longer on Part II than on Part I.

This hypothesis may be employed to explain the findings that normal children do not evidence an increase from Part I to Part II of the game (Zigler et al., 1958), since the assumption can be made that the negative tendency of normal children is much less than that of institutionalized feebleminded subjects. The finding that feebleminded subjects do not show an increase from Part I to Part II of the game when both parts of the game are given under an NS condition (Zigler et al., 1958) may be due to the inadequacy of the NS condition in decreasing the negative tendency. Studies investigating this hypothesized negative reaction tendency are now in progress.

It may be concluded that this study offers evidence that differences among subjects in the frequency with which they manifest certain behaviors which have often been characterized as rigid (perseveration, stereotypy, etc.) are related to differences in the amount of social deprivation the subjects have experienced. It supports the position advanced in earlier studies (Stevenson & Zigler, 1957; Zigler et al., 1958) that constructs used to explain the performance of normal individuals are adequate in explaining the performance of feebleminded ones. This study also offers further evidence of the danger in describing a group as though its members uniformly possess a particular characteristic, e.g., rigidity, without considering such factors as historical differences among subjects, variability in motivation, the nature of the reinforcement provided, or the tasks employed.

SUMMARY

Sixty feebleminded children with MAs of 4 to 8 years played a simple, monotonous game until satiated. The game consisted of two highly similar parts. Following satiation on Part I of the game, the subjects were presented Part II and were again allowed to play until satiated. Two conditions of reinforcement were introduced. Half of the subjects received both verbal and nonverbal support from the experimenter during Part II. The other half of the subjects received only the experimenter's attention during Part I and the experimenter's verbal and nonverbal support during Part II. Two judges read the social histories of the subjects and rated each subject on the degree of pre-institutional social deprivation.

The hypothesis was advanced that interaction with an adult and adult approval provide greater reinforcement for the responses of feebleminded subjects who have experienced a greater amount of social deprivation than for those who have experienced a lesser amount of social deprivation. Three of the four predictions derived from this hypothesis were confirmed. It was found that the more socially deprived subjects spent a greater amount of time on the game, more frequently made the maximum number of responses allowed by the game, and, evidenced a greater increase in time spent on Part II of the game over that spent on Part I. The findings related to the

fourth prediction, that the more socially deprived subjects would make a lower proportion of errors, were in the direction predicted, but failed to reach the .05 level of significance.

These results offer further support for the view that the "rigid" behavior observed in feebleminded subjects may be attributed to a higher motivation to maintain interaction with an adult and to secure approval from him through compliance and persistence.

The assumption that there would be an increase in time spent on Part II of the game over Part I under both conditions of reinforcement received support. However, the two conditions of reinforcement produced no significant differences in the increase of time spent on Part II over that spent on Part I, the time spent on Part I alone, the time spent on Part II alone, nor the total time spent on the game. The further assumption that institutionalized feebleminded subjects come to the experimental situation with both a positive and a negative tendency to respond was advanced to explain the increase in time spent on Part II compared to Part I.

REFERENCES

BOWLBY, J. Maternal care and mental health. *Bull. WHO*, 1951, **3**, 355–533.

CLARK, L. *The nature and treatment of amentia*. Baltimore: Wood, 1933.

GEWIRTZ, J. Social deprivation and dependency: A learning analysis. Paper presented at Dependency in Personality Development Symposium. Amer. Psychol. Ass., New York, 1957.

GEWIRTZ, J., & BAER, D. Deprivation and satiation of social reinforcers as drive conditions. *J. abnorm. soc. Psychol.*, 1958, **57**, 165–172. (a)

GEWIRTZ, J., & BAER, D. The effect of brief social deprivation on behaviors for a social reinforcer. *J. abnorm. soc. Psychol.*, 1958, **56**, 49–56. (b)

GOLDFARB, W. Psychological privation in infancy and subsequent adjustment. *Amer. J. Orthopsychiat.*, 1945, **15**, 247–255.

KOUNIN, J. Experimental studies of rigidity as a function of age and feeblemindedness. Unpublished doctoral dissertation, State University of Iowa, 1939.

KOUNIN, J. Experimental studies of rigidity: I. The measurement of rigidity in normal and feebleminded persons. *Character Pers.*, 1941, **9**, 251–272. (a)

KOUNIN, J. Experimental studies of rigidity: II. The explanatory power of the concept of rigidity as applied to feeblemindedness. *Character Pers.*, 1941, **9**, 273–282. (b)

KOUNIN, J. The meaning of rigidity: A reply to Heinz Werner. *Psychol. Rev.*, 1948, **55**, 157–166.

LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1936.

PLENDERLEITH, M. Discriminaiton learning and discrimination reversal learning in normal and feebleminded children. *J. genet. Psychol.*, 1956, **88**, 107–112.

RIBBLE, M. Infantile experience in relation to personality development. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 621–651.

SARASON, S. *Psychological problems in mental deficiency*. (2nd ed.) New York: Harper, 1953.

SARASON, S., & GLADWIN, T. Psychological and cultural problems in mental subnormality: A review of research. *Genet. psychol. Monogr.*, 1958, **57**, 7–269.

SKEELS, H., UPDEGRAFF, R., WELLMAN, B., & WILLIAMS, H. A study of environmental stimulation, an orphanage preschool project. *U. Ia. Stud. child Welf.*, 1938, **15**, No. 4.

SPITZ, R. A. Hospitalism: An inquiry into the genesis of psychiatric conditions in early childhood. In Anna Freud, H. Hartmann, & E. Kris (Eds.), *The psychoanalytic study of the child*. Vol. 1. New York: International Univer. Press, 1945. Pp. 53–74.

STEVENSON, H., & ZIGLER, E. Discrimination learning and rigidity in normal and feebleminded individuals. *J. Pers.*, 1957, **25**, 699–711.

ZIGLER, E. Preinstitutional social deprivation and rigidity in the performance of feebleminded children. Unpublished doctoral disertation, University of Texas, 1958.

ZIGLER, E., HODGDEN, L., & STEVENSON, H. The effect of support on the performance of normal and feebleminded children. *J. Pers.*, 1958, **26**, 106–122.

(Received March 28, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 422–426

CASE REPORT

MOTIVATIONAL STRUCTURE OF AN HOSPITALIZED MENTAL DEFECTIVE

ANNA M. SHOTWELL
Pacific State Hospital

JOHN R. HURLEY[1] AND RAYMOND B. CATTELL
University of Illinois

FREUD, McDougall, Murray, and others, in attempting to isolate basic dimensions of motivation, did not have the benefit of the factor analytic methods which have enabled Cattell and his co-workers to substantiate the concepts of drive, ego structure, and sentiment. On the assumption that factor analysis of many scores obtained on objective attitude measuring devices would reveal a limited number of motivational components which could account for the multitude of phenomena observed in dynamic behavior, Cattell pioneered three researches: an R-technique study to isolate and identify the basic motivational determinants (Cattell, 1950), another R-technique study to confirm the findings of the first and to develop alternative objective measuring methods (Cattell & Miller, 1952), and a P-technique study to confirm further and to clarify the ideas elicited by the other two studies as well as to investigate the applicability of these ideas to an individual (Cattell & Cross, 1952). The present case study continues this line of research.

The objective here was to discover the motivational dimensions of an institutionalized mental defective by objective P-technique factor analysis and to ascertain whether the findings were congruent with clinical understanding of the case and with earlier factor analytic results on nondefectives.

METHOD

The subject, a female patient at Pacific State Hospital in California, was a heavy-set 15-year-old Negro mental defective of mild degree who had come from an unfit home. Her mother was sexually promiscuous and probably of retarded intelligence. After the subject's father molested her sexually, her mother killed him with the subject a witness to the killing. The subject was described by the hospital staff as quiet, somewhat withdrawn, and undemonstrative, but usually cooperative. For a period of 100 days (5 days per week for 20 weeks), she was given a series of tests and was rated on various traits by her school teachers, by the technician in charge of her ward, and by herself. The test instruments, constructed on the assumption that drives and sentiments are revealed as patterns among attitudes, were designed to measure attitudes in an objective manner by means of ratings, questions about activities, word association, and paired comparisons. Each previously known dynamic factor in the literature was represented by at least two marker variables, i.e., variables which were strongly loaded in the previously known factor were carried over in some form and used in this investigation.

To distinguish each type of evaluative variable from all other variables in the factor matrix, a letter prefix was assigned as follows: RT, ratings by school teacher; RW, ratings by ward technicians; RS, self-ratings; QA, questions regarding activities; WA, word association; and PC, paired comparisons.

The subject was rated by herself and two others on 14 traits. The best possible score was 10; the worst possible was 0. Of the 42 trait ratings, only 9 proved variant enough to be retained for analysis.

Most of the original 22 questions regarding the subject's daily activities required only a "yes" or "no" answer. A few, such as the question, "How many cigarettes have you smoked?" required a quantitative answer. Of the original group of 22, 9 questions were retained for analysis, the remainder being eliminated because they displayed little or no variance.

Thirty-four word association items were presented to the subject in the format:

top. . . . . lid____ winner____

The second and third words of the series were assumed to represent one of several hypothesized drives and sentiments. From these two words, the subject was to choose the one which she first associated with the first word of the series. It was assumed that the choice of one word over another would be determined by her attitudes. The subject's choices were then pooled to obtain a score for each of the hypothesized drives and sentiments. Since the sequence of items was changed from day to day, there was no chance for the subject to mechanize her responses.

Paired comparisons consisted of having the subject indicate which of a pair of statements, each printed on a separate slip of paper, most strongly represented her present desires. Sixteen statements were constructed so as to define each attitude by inclusion of its parts: the stimulus situation, an organism, a need, a course of action, and an object concerned in that course of action. One of the statements was, "I want a good looking boy to make love to me." Since each statement was paired against every other statement, the subject made 120 choices each day for 100 days. Again the sequence of

[1] Presently at System Development Corporation, Paramus, New Jersey.

presentation was altered daily to obviate the possibility of the subject repeating responses mechanically. The 16 statements as well as the Word Association test were markers for seven drives and one sentiment. The seven drives were sex-mating, gregariousness, appeal-dependence, escape-fear, self-assertion, parental-protective, and narcissistic play. The eighth motivational component hypothesized was the self-sentiment.

Analysis of data. As a result of the above procedure, each of the original 98 variables was represented by measures on 100 items. Means, standard deviations, and all possible intercorrelations were computed for each of the 98 variables. Upon inspection of the results, the decision was made to eliminate 56 of the original 98 variables. The major reason for this was that the data displayed little or no variance for these measures. The intercorrelations of the remaining 42 variables were then factor analyzed by the complete simple centroid method (Thurstone, 1947). The diagonal values were set at unity. The resulting factor matrix was then rotated to orthogonal simple structure by the method of quartimax (Neuhaus & Wrigley, 1954). Quartimax failed to disperse fully the main centroid. Consequently, the unrotated factor matrix was again rotated, this time to oblique simple structure. The method used was a combination of oblimax (Pinzka & Saunders, 1954) and Hurley's analytical method.[2] A decent oblique simple structure was finally obtained, with a total of 11 rotated factors remaining for interpretation.[3]

DISCUSSION

Of the 11 factors found, 7 corresponded to factors found by previous researchers. Of 2 new factors, 1 did not lend itself to precise interpretation, and 1 was a subfactor. The factors are shown below in estimated order of variance accounted for.

Factor I—The Sex Erg

		Loading
PC 10	I want a good-looking boy to make love to me	.70
PC 11	I want to go to a good movie every week	.64
RT 1	Lack of shyness	.57
PC 25	I don't want to keep good control of my temper	.53
PC 13	I want to watch games like baseball	.51
PC 20	I don't want to help people in trouble	.50
PC 12	I want to take an active part in recreation	.49
PC 21	I don't want my parents always to have what they need	.39
WA 35	Sex-mating (word association)	.12

This component is one which has accounted for the most variance in all of the studies in which it has appeared. Variables PC 10 and PC 11 label this as a sex component. The other variables, in general, contribute to the picture of an uninhibited, amoral drive which has already found expression in sports and recreation. Variable WA 35 is included, even though the loading is low, in order to show that its direction lines up with the sense of the factor as illustrated by the other variables.

Factor II—Fear-Anxiety Erg

		Loading
PC 17	I want there to be fewer deaths	.54
RT 1	Shyness	.40
PC 16	I want the U. S. to have better protection against the A-bomb	.39
RS 8	Amount of worry (high)	.35
RS 5	Cooperation (low)	.34
PC 23	I don't want time to enjoy my own company	.33
QA 30	Did you watch TV yesterday? (Yes)	.21
WA 38	Escape-fear (word association)	.08

The general sense here is of a motivational component akin to anxiety. Variable RT 1 indicates withdrawal, whereas Variable PC 23 would seem to leave the subject no alternative but extraversion. Thus, the central problem of almost any anxiety syndrome is seen. This factor also removes any doubt as to one of the most useful functions of the television set, the poor man's Miltown.

Factor III—Self-Regard Sentiment

		Loading
RW 3	Evenness of temper	.73
RW 4	Daydreaming	.71
PC 24	I want to do things to make people respect me	.37
PC 13	I don't want to watch games like baseball	.33
PC 11	I don't want to go to a good movie every week	.31
RS 7	Loneliness	.31
PC 25	I want to keep good control of my temper	.24
WA 42	Self-sentiment (word association)	.22

We have here a pensive, self-regarding quality, coupled with a resolve to perform in a socially

[2] The unrotated factor matrix was first rotated by the method of oblimax to give an idea of the oblique structure. Then a rotated matrix V_{ij} was postulated on the basis of the oblimax solution. V_{ij} was constructed to display simple structure. The matrix equation $F_0t = V_{ij}$ was then solved for t, the transformation matrix. All statistical computations were performed on the University of Illinois digital computer, ILLIAC.

[3] Tables containing the unrotated factor matrix, the transformation matrix, the rotated factor matrix, and the cosine matrix have been deposited with the American Documentation Institute. Order Document No. 6742 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

acceptable fashion in one's activities. It is thought of as a sentiment rather than the drive of self-assertion (Cattell, 1957) because its pattern more closely fits that of the previously defined sentiment (Cattell, 1950; Cattell & Cross, 1952; Cattell & Miller, 1952). This sentiment is somewhat analogous to what Horney (1950) means in speaking of the "tyranny of the should."

Factor IV—Gregarious Erg

		Loading
WA 42	Self-sentiment (word association)	.52
WA 36	Gregarious drive (word association)	.49
WA 41	Narcissistic-play drive (word association)	.49
PC 12	I want to take an active part in recreation	.37
PC 13	I want to watch games like baseball	.36
PC 11	I want to go to a good movie every week	.27

"Man is a social animal . . ." is here again restated in factor analytic terms. A drive to be part of the crowd is manifested by the desire for participation in essentially social activities such as baseball and movies. The gregarious erg is not as clear here as in previous research. It seems here to be mixed in with goodly helpings of self-regard and self-love or desire for self-enhancement. Variables WA 36, PC 12, and PC 13 are markers from previous research. This factor demonstrates superb simple structure and has no significantly high correlations with any other factors. Perhaps further research will be able to rotate the confusing Variables WA 42 and WA 41 out of the picture.

Factor V—Appeal-Dependence Erg

		Loading
WA 39	Lack of self-assertion (word association)	.62
WA 38	Lack of escape-fear (word association)	.54
RS 9	Poor state of health	.39
PC 14	I want to feel close to God	.35
PC 15	I want to have the advice of my parents	.34
WA 37	Appeal-dependence drive (word association)	.32

The strong appearance here of Variables PC 14, PC 15, and WA 37 leads to the conclusion that this factor represents the dependency drive. It is the "Mama's boy," the homesick child at camp, and the confused neurotic's plea for help. A lack of self-assertion and escape-fear is entirely consistent with this interpretation, as is a poor state of health. Picture the person who has been "down on his luck" for a relatively long period of time.

Factor VI—Menstruation-Depression

		Loading
PC 10	I don't want a good-looking boy to make love to me	.43
PC 23	I want time to enjoy my own company	.41
RS 9	Poor state of health	.29
QA 34	Are you menstruating? (Yes)	.25

A more solid, impressive display of the efficacy of factor analysis could hardly be found outside the textbooks of Thurstone and Burt. It naturally follows that a girl who is menstruating considers herself in a poor state of health, wants to be alone, and especially wants no amorous entanglements at this delicate time. This type of factor has not appeared before in factor analytic research.

Factor VII—Self-Assertion

		Loading
PC 18	I want to earn more money	.52
PC 15	I don't want to have the advice of my parents	.48
QA 27	Did you get a letter from home yesterday?	.40
WA 40	Parental-protective drive (word association)	.38
QA 28	Did you write a letter home yesterday?	.28
PC 23	I don't want time to enjoy my own company	.27
RT 2	Reduced quantity of work	.26
RS 7	Loneliness	.25
PC 19	I want to succeed in my life's work	.25
PC 22	I don't want to spend more time on smoking or drinking	.24
RS 6	Happiness	.24
WA 39	Self-assertion drive (word association)	.07

This factor has the poorest simple structure in the rotated reference vector matrix. The variance of the factor is spread thin over many seemingly unrelated variables. The presence of Variables PC 18, PC 15, PC 19, and PC 22 lead strongly to the interpretation that this represents a motivational component of self-assertion even though the factor is generally muddy. Perhaps further research will provide the answer to the unsatisfactory pattern found here.

Factor VIII—Sub-Factor: "Good Time Yesterday"

		Loading
QA 29	Did you spend any money yesterday?	.54
QA 31	Did you read any books on the ward yesterday? (Yes)	.49
QA 32	Did you eat between meals yesterday? (Yes)	.46
QA 33	Did you play any games on the ward yesterday? (Yes)	.46
PC 22	I want to spend more time on smoking and drinking	.26
PC 15	I don't want to have the advice of my parents	.22

This factor correlates highly with Factor IX, Narcissism (.51), and has no real substance in itself. Therefore, it is interpreted as a subfactor of Factor IX. This interpretation means that in another experiment, with possibly a larger number of measurements being taken, this "factor" would resolve itself into Factor IX.

Factor IX—Narcissism Erg

		Loading
QA 31	Did you read any books on the ward yesterday? (Yes)	.68
QA 33	Did you play any games on the ward yesterday? (Yes)	.58
PC 22	I want to spend more time on smoking or drinking	.38
QA 27	Did you get a letter from home yesterday? (No)	.32
RS 8	Low amount of worry	.32
PC 23	I want time to enjoy my own company	.28
QA 26	Did you dream about your boyfriend last night? (Yes)	.28
WA 41	Narcissistic-play drive (word association)	.14

As previously mentioned, most of the variables of Subfactor VII are included here as a subset. This component is interpreted as an erg of narcissism, a self-love or desire to make oneself feel better, etc. In Freudian terms, it represents one of the paths along which the self-worshipping libidinal energy may flow. It compels its possessor to have interest investments in sex, drinking, and other "fun" activities. In the realm of clinical diagnosis, the clinician would suspect conflict if this factor and Factor III, the self-regard sentiment, were simultaneously running at relatively high levels in the same person.

Factor X

		Loading
RS 7	Low amount of loneliness	.45
PC 22	I want to spend more time on smoking or drinking	.36
RS 6	Unhappiness	.34
RS 8	I don't want to succeed in my life's work	.32
PC 14	Low amount of worry	.31
PC 19	I don't want to feel close to God	.30
PC 20	I want to help people in trouble	.30

There is seemingly no plausible interpretation which can be made of Factor X. It is reproduced here as much as an indication of the writers' bewilderment as for the reader's amusement. An awkward, tentative hypothesis is that this is a component somehow connected with philanthropy, a sentiment possessed by one with much money (no worry) and many friends (no loneliness) but with an urge to help others. Variables RS 8 and PC 19 may indicate guilt feelings.

Factor XI—Repressed Sex Component

		Loading
WA 35	Sex-mating (word association)	.39
WA 40	Parental-protective drive (word association) (negative)	.35
PC 23	I want time to enjoy my own company	.34
QA 26	Did you dream about your boyfriend last night? (Yes)	.31
QA 27	Did you get a letter from home yesterday? (No)	.31

This is the last factor to be retained from the factor matrix. As such, the variance which it accounts for is necessarily lower than that accounted for by the other 10 factors. It is here interpreted as a component of repressed sex energy. The clash of feelings for one's parents and feelings for oneself comes into sharp focus in this pattern. The rise of this factor in intensity from day to day could be a signal of dammed-up sex energy looking for release. The sense of this factor, even though it correlates .37 with the Sex Erg, is not one of a constant flow of sexual vitality. Rather, it suggests the accumulation of sex vitality which has not yet gained expression.

SUMMARY

Objective tests whose motif was structured after that of devices employed by previous researchers were administered to a maladjusted mental defective who was a patient in a mental institution. The scores on the tests were correlated and factor analyzed. On interpretations of the factors after appropriate rotation to oblique positions, the factors so discovered appeared to be essentially congruent with previous research. Further, the factor pattern that emerged was capable of being interpreted as the motivational structure of the subject.

The possibility that not all ergs (drives) and metanergs (sentiments) were found does not detract from the basic fact that with suitable instruments one can measure and reveal the fundamental dynamic structures of a maladjusted mental defective. The implications of this for clinical practice are promising. Further confirmatory research is needed to furnish a valid, reliable instrument which can be successfully and economically used as both a diagnostic and a predictive tool. The hope that such a tool can be produced depends on the degree to which clinicians and psychometrists can work together.

REFERENCES

CATTELL, R. B. The discovery of ergic structure in man in terms of common attitudes. *J. aborm. soc. Psychol.*, 1950, **45**, 598–618.

CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.

CATTELL, R. B., & CROSS, K. P. Comparison of the ergic and self-sentiment structures found in dynnamic traits by R- and P-techniques. *J. Pers.*, 1952, **21**, 250–271.

CATTELL, R. B., & MILLER, A. A confirmation of the ergic and self-sentiment patterns among dynamic traits by R-technique. *Brit. J. Psychol.*, 1952, **43**, 280–294.

HORNEY, KAREN. *Neurosis and human growth.* New York: Norton, 1950.

NEUHAUS, J. O., & WRIGLEY, C. The quartimax method. *Brit. J. statist. Psychol.*, 1954, **7**, 81–92.

PINZKA, C., & SAUNDERS, D. R. Analytic rotation to simple structure: II. Extension to an oblique solution. (RB-54-31) Educational Testing Service Bulletin, 1954. (Multilithed)

THURSTONE, L. L. *Multiple-factor analysis.* Chicago: Univer. Chicago Press, 1947.

(Received January 15, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 427–430

CRITIQUE AND NOTES

GSR CONDITIONING, ANXIETY, AND EXTRAVERSION

WESLEY C. BECKER AND HALSEY H. MATTESON
University of Illinois

USING basic notions derived at least in part from different aspects of Hullian theory, Eysenck and Spence have developed two different *but not necessarily mutually exclusive* hypotheses concerning the relationship of personality characteristics to conditioning. Eysenck (1955) has proposed that individuals in whom reactive inhibition develops rapidly and dissipates slowly tend to develop extraverted behavior patterns. Since reactive inhibition interferes with the repetition of a conditioned response, Eysenck has predicted that extraverts will condition less readily than introverts. On the other hand, Spence (1958) has hypothesized that the higher emotional reactivity of the anxious person acts like other drives in increasing reaction potential during conditioning; therefore, a positive relationship can be predicted between the degree of manifest anxiety and level of conditioning.

Studies from Eysenck's laboratory have supported Eysenck, whereas studies from Spence's laboratory have supported Spence. Franks (1956) found that dysthymics (introverted neurotics) conditioned at a higher level than hysterics (extraverted neurotics), using the GSR and eyeblink conditioning simultaneously. In a second study, using normal subjects, Franks (1957) found a significant relationship between the Extraversion (E) scale of the Maudsley Personality Inventory and the level of eyeblink conditioning. No significant relationship was found between the Neuroticism (N) scale of the Maudsley Personality Inventory and conditioning. Since N is highly correlated (approximately .90) with the Taylor Manifest Anxiety scale (*MA*) used by Spence, Franks argued that his data gave no support to Spence's theory.

At the Iowa laboratory, Farber, Spence, and Bechtoldt (1957) found eyeblink conditioning positively related to the *MA* but not to Guilford's Rhathymia (R) scale. R is highly correlated with E and has also been used by Eysenck as a measure of extraversion. Several additional studies showing significant relationships between eyeblink conditioning and manifest anxiety were summarized by Spence (1958).

The impetus for the present study came from a recent failure by Becker (1959) to find support for either Eysenck's or Spence's hypothesis when GSR conditioning was used on *unselected* male and female college students. In this study, Becker found excellent evidence that the measures were reliable and reflected a conditioning process, but no relationships were found to questionnaire measures of introversion or anxiety. While both Eysenck[1] and Spence (1958) have been unhappy with attempts to test their hypotheses using GSR conditioning, each of them can point to one study with the GSR which suggests support for their theories. Franks' (1956) simultaneous use of GSR and eyeblink conditioning has already been mentioned as supporting Eysenck's hypothesis. In his recent review, Spence (1958) reported that subjects selected on the basis of greater GSR and heart-rate changes to noxious stimuli showed more eyeblink conditioning. These results suggested a re-examination of the relationship of anxiety and extraversion to GSR conditioning, using extreme anxiety and extraversion groups (to compensate for unreliability) and all male subjects (to rule out complications from suspected sex differences).

METHOD

Guilford's R scale and Cattell's Anxiety (A) scale were given to 273 male college students (nonvolunteer) enrolled in introductory psychology. Eysenck previously used R as a criterion of extraversion, and A is known to be highly correlated[2] with other measures of anxiety. Four groups of 10 subjects with extreme scores on both A and R were selected for conditioning. The cut-off scores for these four extreme groups were determined from a scatter plot so that the 10 subjects in each group would deviate from the total means for A and R by approximately the same amount.

A Hunter Model 300 GSR amplifier with visual reading was used to measure basal resistance and resistance changes as a result of stimulation. A wire wound potentiometer was used to control shock level and a Pentron tape recorder for presentation of stimuli.

After acquainting the subject with the operation of the apparatus, finger electrodes were attached to the index and third fingers of the left hand, and shock electrodes were attached to the tender side of the right arm just above the elbow. All contact areas were cleaned with alcohol and rubbed with electrode jelly. Care was taken to avoid circulation blockage in attaching

[1] Personal communication.

[2] With 80 college males, Becker (1959) has found correlations of .79 with Eysenck's N, and .79 and .83 with Guilford's C and D, respectively. All three of these measures correlate about .85 with the Taylor *MA* (Farber et al., 1957).

electrodes. The GSR was calibrated and the maximum shock level the subject would tolerate was determined. Then, to allow time for adaptation to the situation, the experimenter played a card identification game with the subject. All the above took 10 minutes. After being given a catalogue of courses to look over, the subject was instructed as follows: "I am going to play a tape with random words on it. You do not need to pay attention to the words, so you can just read along." A list of 104 words spaced by 5-second intervals was then played twice. The word "repeat" occurred 20 times on this list and was followed immediately by a 1-second shock on 12 of the trials. Eight of the repeat trials were considered conditioning test trials. The word crafty preceded repeat 12 times, thus permitting a measure of temporal generalization, and the word receipt occurred 7 times randomly in the list, thus permitting a measure of generalization on the basis of sound similarity. Finally the word constant appeared 11 times in the list, but never in a space just preceding the shocked word. This served as a control on the effect of frequency of word presentation.

Shock was administered by a hand switch. When the GSR to the shock given by any subject decreased to a noticeable degree over trials, the shock level was increased. This was a crude attempt to maintain equivalence of the UCS for all subjects and to compensate for such factors as adaptation to shock, loosening of the shock electrodes, and the like. Basal resistance was adjusted before each word when necessary to center the meter. Recordings of basal were made at the beginning, in the middle, and at the end of both series of trials. The maximum decrease in resistance to each stimulus word was recorded.

The problem of determining appropriate measures of conditioning is a critical one. Two scores were used. The first is simply the mean conditioned response amplitude. The second is a criterion-type score based on the number of trials on which the conditioned response was one-half or greater than the preceding unconditioned response. Conditioned response amplitude is clearly a measure of reaction potential ($_sE_R$) and thus is appropriate in testing Spence's theory. Hull (1943) hypothesized that GSR amplitude is a direct linear function of $_sE_R$. The criterion measure, on the other hand, eliminates sources of variation in the conditioning score which could not be logically a function of reactive inhibition. The greatest such source might be called general autonomic reactivity, which should be related to Hull's D but not I_R. In a previous study Becker (1959) found thet CR amplitude correlated .67 and .78 with UR amplitude in two different testing sessions. These correlations and similar findings by others of a general autonomic responsiveness factor indicate a need to control for such variance in the conditioning measure when testing Eysenck's hypothesis. Becker (1959) found essentially zero correlations between the criterion conditioning score and UR amplitude, indicating that this source of variation is successfully controlled by the criterion score. The criterion measure also controls to some degree for adaptation effects, basal resistance artifact, and differential subject shock levels to the degree that these factors affect the CR and UR equally. Logically, the criterion conditioning measure makes sense in that conditioning is considered stronger the closer CR approximates UR.

Responses to generalization stimuli provide still another kind of measure of conditioning. However, there are many theoretical ambiguities concerning the effects of both reactive inhibition and drive level on generalization so that clear-cut predictions could not be made. The generalization data will be examined solely for their empirical interest.

RESULTS

In the larger sample of 273 subjects, A and R were found to correlate −.13. For the 40 subjects in the conditioning sample, the correlation between A and R was −.06. Split-half reliabilities for A and R, respectively, were found to be .76 and .83 on 273 students. Thus it is unlikely that differences in magnitude of relationship of the conditioning measures to A and R can be attributed to unreliability in the personality measures.

The odd-even reliability of the criterion conditioning score was .68 and that for the CR amplitude score .93. Using similar procedures, Becker (1959) found for the criterion conditioning score a split-half reliability of .80 and a between-sessions reliability of .71. For the CR amplitude score, Becker (1959) found a split-half reliability of .84 and a between-sessions reliability of .63.

A critical problem facing the experimenter using the GSR in a conditioning study is to demonstrate that conditioning has actually taken place. Since GSR measures tend to show adaptation effects over time, the usual learning curve cannot be used as evidence. However, it is possible to compare the subjects' responses to the conditioned stimulus and the generalization stimuli with their responses to neutral stimuli. Table 1 presents a comparison for both the amplitude scores and a criterion-type score based on a ratio of magnitudes rather than frequency. This latter measure correlates over .90 with the criterion measure used in the main analysis, but it illustrates more clearly the conditioning effect. From Table 1 it is apparent that significant learning effects are reflected in the conditioning and generalization measures.

The summary of mean amplitude scores by personality groups is presented in Table 2. Since the mean basal resistance measure correlated .35 with the conditioned response amplitude measure and there were group differences in basal resistance, it was considered necessary to correct for the differences in basal resistance by the use of covariance analysis. As would be expected, the high anxiety (HA) subjects tended to have lower basal resistances. This lower basal level in the HA groups would tend to obscure the predicted positive relationship between A and GSR amplitude. Table 3 presents the group means after covariance ad-

TABLE 1
EVIDENCE FOR CONDITIONING

Stimulus	Amplitude Scores		Ratio of Indicated Stimulus to Unconditioned Stimulus	
	M	σ	*M*	σ
Unconditioned Stimulus	14.8	7.3		
Conditioned Stimulus	6.7*	4.3	.45*	.19
Generalization Stimuli	3.8*	2.3	.26*	.15
Control Word "Constant"	2.4	1.8	.16	.16
All Other Words	2.0	1.5	.14	.09

*Significantly different from control words beyond the .001 level (*t*'s ranged from 7.5 to 11.3).

TABLE 2
MEAN AMPLITUDE SCORES UNADJUSTED

	Low R	High R	Row *M*
HA	8.1	8.1	8.10
LA	5.3	5.1	5.35
Column *M*	6.70	6.75	

TABLE 3
MEAN CR AMPLITUDE SCORES, COVARIANCE ADJUSTED FOR BASAL RESISTANCE DIFFERENCES

	Low R	High R	Row *M*
HA	8.8	8.3	8.55
LA	5.3	4.4	4.85
Column *M*	7.05	6.35	

TABLE 4
MEAN CRITERION CONDITIONING SCORES

	Low R	High R	Row *M*
HA	8.7	8.8	8.8
LA	7.0	6.3	6.7
Column *M*	7.9	7.6	

justment for basal differences. The analysis of covariance gave an *F* ratio of 8.44, significant at the .01 level, for the difference in GSR amplitude between the HA and LA groups. No other effect approached significance. These findings support Spence's theory.

Group means for the analysis of the criterion conditioning scores are presented in Table 4. The only effect to approach statistical significance was the difference between HA and LA groups ($F = 2.82$, $p < .10$). The findings give no support for Eysenck's theory. The generalization scores showed no significant differences between groups.

DISCUSSION

The results fail to support Eysenck's theory while giving support to Spence's notions. Before turning to the implications of these results, brief mention should be made of a problem of control typical of most shock studies. Lacking a reliable measure of shock current through the subject, an attempt was made to equalize the uncontrolled factors of skin resistance, adequacy of contact, and the like, by requiring each subject to receive the maximum shock level he could tolerate. While systematic study demands that better methods of control of shock level be developed, it is possible to assess the possible effects of lack of such control on the present findings by examining the mean shock levels by groups and the correlations between shock level and conditioning scores. Examination of group means reveals no significant differences between high and low groups. Nevertheless, the mean shock level for the high R groups is still considerably higher than that for the low R groups (246 volts versus 218 with σ of 47 and 66, respectively). If shock level were positively correlated with the criterion conditioning score used to test Eysenck's hypothesis, then this difference in shock level could have acted to suppress a real relationship between GSR conditioning and R. However, the correlation between shock level and the criterion conditioning score was found to be −.11. The correlation between shock level and the amplitude conditioning score was −.29. The negative relationships, though insignificant, may be related to the fact that shock level was increased if the GSR to shock decreased appreciably. The above findings support the conclusion that the method of determining shock level did not produce systematic biasing effects which would necessitate modification of the main findings.

In considering the implications of this study for Eysenck's inhibition theory, the reader should recognize the possibility that the GSR response mechanism functions according to completely different principles than a skeletal motor response. The eyeblink response serves as a protective avoidance response, while the GSR to shock is noninstrumental and probably adaptively superfluous. The present findings may be interpreted as only raising a question about the applicability of Eysenck's theory to nonskeletal responses. However, this report is an addition to that growing list of failure by American psychologists to confirm Eysenck's reactive inhibition theory of extraversion (Becker, 1959; Farber et al., 1957; Miller, 1957; Ray, 1959; Rechtschaffen, 1958).

SUMMARY

Hypotheses by Eysenck and Spence relating GSR conditioning to extraversion and anxiety, respectively, were examined. Four groups of 10 subjects scoring on the extremes of anxiety and extraversion measures were selected from a group of 273 male introductory psychology students for conditioning. Using a conditioned response amplitude measure, which appeared to be the more appropriate measure for testing Spence's hypothesis, a significant positive relationship was found between anxiety and conditioning. No significant relationship was found between extraversion and conditioning when either the conditioned response amplitude measure was used or (more appropriately) when a criterion-type conditioning measure was used. The results give clear support to Spence's theory while failing to support that of Eysenck.

REFERENCES

BECKER, W. C. Cortical inhibition and extraversion-introversion. Unpublished report, University of Illinois, 1959, USPHS Grant M-1713. (Mimeo)

EYSENCK, H. J. Cortical inhibition, figural aftereffect, and theory of personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94–106.

FARBER, I. E., SPENCE, K. W., & BECHTOLDT, H. P. Emotionality, introversion-extraversion, and conditioning. Paper presented at the Midwestern Psychological Association, Chicago, May 1957.

FRANKS, C. M. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143–149.

FRANKS, C. M. Personality factors and the rate of conditioning. *Brit. J. Psychol.*, 1957, **48**, 119–126.

HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.

MILLER, PAULA E. Reactive inhibition in extraversion and introversion. Unpublished bachelor's thesis. University of Illinois, 1957.

RAY, O. S. Personality factors in motor learning and reminiscence. *J. abnorm. soc. Psychol.*, 1959, **59**, 199–203.

RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283–291.

SPENCE, K. W. A theory of emotionally based drive (*D*) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

(Received January 14, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 431-432

RELATIONSHIP OF PERSUASIVE INTERACTION TO CHANGE OF OPINION IN DYADIC GROUPS WHEN THE ORIGINAL OPINIONS OF PARTICIPANTS ARE EXPRESSED PRIVATELY AND PUBLICLY

VLADIMIR B. CERVIN
Imperial Oil Limited

ROBERT C. JOYNER, J. MICHAEL SPENCE, AND RUDOLPH HEINZL
University of Toronto

RECENT research on the relationship between neurotic anxiety symptoms and persuasibility has yielded contradictory results. Some studies (Cervin, 1956, 1957b; Janis, 1954) indicate that low persuasibility is associated with high anxiety, others (Janis & Hovland, 1959) fail to confirm this finding.

Other research (Hovland, Janis, & Kelley, 1953, 1957) indicates that public commitment to an opinion either tends to freeze the subjects in their opinions and make them less susceptible to persuasive communication, or at least makes the subjects less susceptible to persuasive communication than private commitment (Bennett, 1955). It is possible that commitment has something to do with instability of the relation between personality factors and persuasibility.

The present study is concerned with the relationship of emotional responsiveness (Cervin, 1957a) to persuasibility under conditions of public commitment of subjects to their opinions before the experiment. It is identical to the previous study (Cervin, 1957b) except that in the latter subjects did not commit themselves publicly to their opinions, whereas in the present one they did. Any difference in the results should therefore be attributable to the changed condition of commitment.

PROCEDURE AND RESULTS

Twenty-eight pairs of subjects were selected from among the students of the General Arts Course at the University of Toronto. Within each experimental pair subjects were equated on a number of variables, but differed on scores of emotional responsiveness (Cervin, 1957a). Twenty-two control pairs were equated also on emotional responsiveness. As in the previous study, all subjects filled out an opinionnaire on a number of topics (such as, "Undergraduates should not marry," "Sexual deviates should be institutionalized for life," etc.) of some interest to students. Religious and political topics were eliminated as too emotionally involving. In the experimental situation a topic was selected for each pair on which subjects within the pair disagreed. Their opinions were announced by the experimenter and confirmed by the subjects. Each pair was then instructed to discuss the topic and arrive at a common statement of opinion (agree, disagree, compromise). After the discussion, subjects again recorded their opinions on the topic discussed. If only one subject changed his opinion, this was called one-sided change or persuasion; if both changed, it was called two-sided change or compromise; if neither changed, i.e., subjects agreed to disagree, the case was classified as equivalent to compromise. As before, the analysis was made in terms of the number of pairs in each of these categories, both for experimental and control pairs.

Persuasion. In order to test for relative frequency of one-sided persuasion as compared to compromises, the frequency of one-sided changes of opinion was compared with that of two-sided changes in experimental and in control pairs. The ratios of the number of one-sided to two-sided changes of opinion are 13/15 for experimental, and 10/12 for control pairs. These ratios are not significantly different.

The conclusion therefore is that when opinions are held publicly, the correlation between the difference in the level of emotionality within pairs and the frequency of one-sided changes of opinion is zero, i.e., one-sided persuasion does not occur more frequently in pairs characterized by differential emotionality than in those of equal emotionality. This result is contrary to the result for privately held opinions obtained before, but agrees with that of Janis and Hovland (1959).

Persuasibility. Persuasibility was defined as responsiveness (in terms of change of opinion) to constant communication stimuli, and therefore was measured by the number of pairs with one-sided changes of opinion (persuasion) where communication stimuli were constant. In order to test for persuasibility of high as compared with low emotional subjects, the frequency of changes of opinion by high emotional subjects in experimental pairs that showed one-sided opinion switches was compared with that of low emotional subjects. This ratio is 10/3, i.e., high emotional

subjects changed their opinions in 10 pairs and low emotional subjects only in 3 pairs. The expectation for this ratio, based on the results of the previous experiment, is 3/10. The probability of obtaining 10 or more successes out of 13 in a binomial distribution with $p = 3/13 \cong .23$ and $q \cong .77$, i.e., for $(.23 + .77)^{13}$, is $p < .001$. The conclusion therefore is that when opinions are held publicly, the high emotional subjects tend to change their opinions more frequently than low emotional subjects, i.e., under public commitment persuasibility of high emotional subjects is higher than that of low emotional subjects. This result is again contrary to the result for privately held opinions (Cervin, 1957b).

It may be that the model of persuasion proposed earlier (Cervin, 1957b) in its present form does not apply under conditions of publicly held opinions, or else, under these conditions, a drive reduction occurs in high emotional subjects which brings it down under the drive level of low emotional subjects.

REFERENCES

BENNETT, E. B. Discussion, decision, commitment, and consensus in "group decision." *Hum. Relat.*, 1955, **8**, 251–273.

CERVIN, V. Individual behavior in social situations: Its relation to anxiety, neuroticism, and group solidarity. *J. exp. Psychol.*, 1956, **51**, 161–168.

CERVIN, V. Personality dimensions of emotional responsiveness and rigidity, and scales for measuring them. *J. Pers.*, 1957, **25**, 626–642. (a)

CERVIN, V. Relationship of ascendant-submissive behavior in dyadic groups of human subjects to their emotional responsiveness. *J. abnorm. soc. Psychol.*, 1957, **54**, 241–249. (b)

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *Communication and persuasion.* New Haven: Yale Univer. Press, 1953. P. 127.

HOVLAND, C. I., JANIS, I. L., & KELLEY, H. H. *The order of presentation in persuasion.* New Haven: Yale Univer. Press, 1957. P. 131.

JANIS, I. L. Personality correlates of susceptibility to persuasion. *J. Pers.*, 1954, **22**, 504–518.

JANIS, I. L., & HOVLAND, C. I. *Personality and persuasibility.* New Haven: Yale Univer. Press, 1959. P. 61.

(Received November 16, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 433–434

PERSONAL CONSTRUCTS AND THE SEMANTIC DIFFERENTIAL[1]

SPIRO B. MITSOS[2]
Evansville State Hospital

KELLY (1955), in his personal construct theory, stresses the importance of each individual's personal dimensions with which he construes events in his world. These dimensions, or personal constructs, are viewed as bipolar dichotomous abstractions along which the person ascribes meaning to his world. All of a person's thinking, problem solving, and predictive behavior are directly related to the nature of his construct system.

This emphasis on personally meaningful constructs, or dimensions of thinking, may have important implications for psychological measurement. Kelly, of course, developed a technique of personality assessment, the Role Construct Repertory Test, that directly follows from his theory of behavior. In introducing his theory and technique to the clinical setting he says the following with regard to tests and testing:

> Whose yardstick does this test represent? When a psychologist attempts to understand a client, he is limited, of course, to those yardsticks which he has accumulated in his own repertory. Perhaps it is more accurate to say that he measures the client within his own set of coordinate axes. The so-called "objective" tests are designed exclusively along the lines of psychologists' axes. The client is plotted with respect to these given axes and his own set of axes is ignored entirely (Kelly, 1955, p. 206).

The semantic differential as developed by Osgood, Suci, and Tannenbaum (1957) involves assumptions quite similar to those of role construct technique. Bipolar construing dimensions—personal constructs in Kelly's system and scales in the semantic differential—are put to quite similar use. But the techniques differ in one potentially very important area. Kelly emphasizes the personal nature of the dimensions and therefore allows each subject to use his own constructs in the examination. With the semantic differential, however, all subjects construe the objects under question with the same set of scales—a set determined by the examiner.

The present study was addressed to the applicability of a personal construct approach within the conventional framework of the semantic differential. The problem was to demonstrate that using personally meaningful dimensions can enhance the semantic differential technique. It was predicted that the introduction of subject oriented personal constructs within the conventional semantic differential should yield concepts more saturated with meaning, without significant change in the total semantic structure.

METHOD

Subjects for this experiment included 16 students in an undergraduate psychology class. Each subject was given a list of seven bipolar adjective scales (Osgood et al., 1957) with demonstrated high factor loadings on the semantic differential "evaluative" factor. The subjects were asked to think about these seven dimensions and select the three that they considered most personally meaningful in thinking about people. Each subject was then presented with seven additional scales loaded heavily on the "potency" factor. Subjects were again asked to select the three most personally meaningful. The same procedure was followed with seven scales with loadings on the "activity" factor. Thus, from 21 scales each subject selected 9 that he felt he could meaningfully use in construing people. Furthermore, these 9 scales were evenly distributed among the three major semantic differential factors. The nature of these factors can be determined by inspection of Table 1 where all scales are listed by factor.

After selecting the scales, each subject followed the procedure of the semantic differential and applied all 21 of the scales to each of the following seven concepts: big business leader, junior executive, foreman, unskilled laborer, politician, labor racketeer, and communist. The ratings were completed in the conventional manner as described by Osgood.

The data were analyzed via Osgood's generalized distance formula, yielding a score (D) which is a measure of the degree to which concepts are similar or different in meaning. In terms of the formula, a D score matrix can be computed with measures of the distance in meaning of each concept from every other concept. A hypothetical concept located at the neutral point of every scale can serve as a norm of meaninglessness, or point of origin, and D scores can be computed indicating the distance of each concept from such a meaningless concept. Concepts furthest from this point of origin are most saturated with meaning. In the D score matrix there is a direct relationship between the magnitude of the score and the degree of distance in meaning between any two concepts.

[1] This investigation was supported in part by a research grant, MY 2244 (C1), from the National Advisory Mental Health Council, National Institutes of Health, United States Public Health Service.

Part of this paper was read at the Spring 1959 meeting of the Indiana Psychological Association.

[2] Now at the Rehabilitation Center, Evansville, Indiana.

RESULTS

Three D score matrices were computed with scale scores grouped by factors. One matrix was based on the factor scores for all 21 scales, the conventional approach to the semantic differential. A second matrix was based on factor scores computed from the nine personal scales selected by the various subjects. The third matrix was computed from factor scores derived from the scale ratings on the three scales least often selected by the 16 subjects for each factor. A list of the 21 scales, together with the frequency tabulations, is presented in Table 1.

TABLE 1

FREQUENCY OF SELECTION OF SCALES AS AMONG THREE MOST MEANINGFUL

Scales	Frequency of Selection ($N = 16$)
Evaluative Factor	
Good–Bad	7
Pleasant–Unpleasant	10
Bitter–Sweet	4
Nice–Awful	6
Fragrant–Foul	0
Honest–Dishonest	11
Fair–Unfair	10
Potency Factor	
Large–Small	11
Hard–Soft	4
Strong–Weak	11
Deep–Shallow	8
Heavy–Light	5
Thick–Thin	2
Rugged–Delicate	7
Activity Factor	
Calm–Agitated	11
Sharp–Dull	14
Ferocious–Peaceful	0
Hot–Cold	5
Angular–Rounded	2
Active–Passive	10
Fast–Slow	6

TABLE 2

DISTANCE MEASURES (D) COMPARING DISTANCES OF CONCEPTS FROM THE POINTS OF ORIGIN IN THREE D SCORE MATRICES

Concepts	Twenty-one Scale Matrix	Personal Scale Matrix	Infrequent Scale Matrix
Big Business Leader	1.91	2.47	.95
Junior Executive	1.67	2.37	.72
Foreman	1.71	2.21	.63
Unskilled Laborer	1.18	1.27	1.07
Politician	1.27	1.29	.98
Labor Racketeer	2.06	2.58	1.63
Communist	1.39	1.92	1.00

Observation of the three D score matrices revealed marked differences in score magnitudes. Scores in the personal scale matrix were substantially larger. The D scores representing the distance of each concept from the point of origin, or the point of meaninglessness, in the semantic field are presented in Table 2. These scores reflect the degree of saturation.

The Wilcoxon matched-pairs signed-rank test was applied to the scores reported in Table 2. The distances from the point of origin in the personally meaningful matrix were found to be significantly greater than the distances based on all 21 scales ($p < .01$). These later distances were in turn significantly greater than those based on the infrequently selected scales ($p < .01$). Significance levels were based on a one-tailed test.

A question is raised as to whether the demonstrated increase in the meaningfulness of concepts resulting from the use of a limited number of personally meaningful dimensions occurs at the expense of "distortion," or changes in the semantic field. Significant changes do not appear to have occurred, since models constructed from the three matrices have very similar shapes, although differing greatly in size.

SUMMARY

In accord with Kelly's (1955) emphasis on the importance of personal meaningfulness to the subject in psychological measurement, the possibility of employing personal constructs within the semantic differential technique was investigated. A direct relationship was predicted between the personal meaningfulness of scales and the degree to which the concepts rated by the scales are saturated with meaning. The prediction was supported. Furthermore, increased saturation does not appear to occur at the expense of "distortion" in the semantic field.

REFERENCES

KELLY, G. A. *The psychology of personal constructs.* New York: Norton, 1955.

OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning.* Urbana: Univer. Illinois, 1957.

(Received November 16, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 435–438

INCONGRUENCY AS A PREDICTOR OF RESPONSE TO HUMOR[1]

DONN BYRNE
University of Texas

JAMES TERRILL
Mental Research Institute, Palo Alto, California

AND PAUL McREYNOLDS[2]
Veterans Administration Hospital, Palo Alto, California

THE investigation of both humor preferences and humor interpretations has a long psychological history, but our ability to predict either of these behaviors is not very impressive as yet. The present study is concerned with the possible relationships between these two aspects of response to humor and a measure of incongruency.

The degree to which a humorous stimulus is amusing to an individual has been found to be in part a function of the content of the humor relative to other measured aspects of the respondent's behavior. For example, humor with hostile content is most appreciated by individuals who express aggressive sentiments on paper-and-pencil tests (Murray, 1934), who are aggressive in a hospital ward setting (Byrne, 1956), and who produce TAT stories with aggressive content (Grziwok & Scodel, 1956). It has been proposed frequently that conflict in a drive area may lead to either a positive or a negative response to related humor (Ghosh, 1939; Levine, 1956; Redlich, Levine, & Sohler, 1951). This suggests a curvilinear relationship between humor preference and conflict.

A second humor variable concerns the ability of individuals to interpret the point being made by the humorist, regardless of their response to the funniness of the humor. Clinical evidence has been reported (Redlich et al., 1951) that failure to understand or distortion of the meaning of cartoons is closely related to personality dynamics. It has been found (Byrne, 1956) that neuropsychiatric patients who express hostility in their everyday behavior are better able to recognize the hostile content of cartoons than patients who fail to express hostility. Senf, Huston, and Cohen (1956) found that schizophrenics had an impaired ability to interpret cartoons, though their study did not include the relating of particular conflicts to specific misinterpretations.

It seems reasonable to suppose that both the extent to which a humor stimulus is perceived as funny and the interpretation given to its content are influenced by the degree of conflict in the drive area represented by the stimulus.

McReynolds (1956) has presented a theoretical approach to the problem of conflict. This conceptualization assumes that percepts representing one's experiences tend to be organized or assimilated into one's apperceptive mass, and that assimilation may be hindered if there are marked *incongruencies* among the conceptual schemata according to which the percepts would be assimilated. By *incongruency* is meant a state of contradiction, disharmony, or inconsistency. Percepts which occur in areas of marked incongruency would therefore tend to remain unassimilated, and the conceptualization proposes that the greater the quantity of unassimilated material, the greater the feeling of anxiety. It holds further that individuals tend to avoid the perception of material which would be difficult to assimilate. From this it can be predicted that subjects having marked incongruency in a given drive area would tend to avoid percepts in this area.

The present study investigated the relationships between (*a*) incongruency and humor preferences in the areas of sex and hostility, and (*b*) incongruency and humor interpretations in the same areas. With regard to the former, a curvilinear relationship was predicted; for the latter, it was hypothesized that the greater the incongruency in a given drive area, the greater is the tendency to interpret cartoons in such a manner as to avoid percepts in that area.

METHOD

Subjects. The subjects consisted of 33 male neuropsychiatric patients at the Palo Alto Veterans Administration Hospital.[3] They were mostly closed ward

[1] From the Research Laboratory, Veterans Administration Hospital, Palo Alto, where all three authors were affiliated when this study was carried out.

[2] The authors wish to express their appreciation to Itsuye Sakai and Gershon Berman for their help on this study.

[3] Subjects were selected from patients who had taken the MMPI in such a manner that no member of the sample had an *F* score of over 70. This procedure

patients, and represented a considerable variety of diagnostic classifications. The median age was 36, with a range from 21 to 65.

Incongruency measures. The general technique has been described elsewhere (McReynolds, 1958). In the present instance, the subject responded twice to each of 60 statements. Twenty of the items were sexual in content (e.g., making love, daydreaming about sex), 20 items had hostile content (e.g., playing a joke on a bully, getting even with someone who has hurt me), and 20 were neutral or buffer items. First, the subject was asked to indicate his *liking* for each of the 60 activities, presented in a dittoed list, by checking one of the following: Mainly Like, Mainly Dislike, Both Like and Dislike, and Neither Like Nor Dislike. Later in the testing session, the subject received the same items in a different order, typed individually on cards. This time the subject's task was to sort the activities according to his *personal moral judgment* into the categories Mainly Good, Mainly Bad, Both Good and Bad, and Neither Good Nor Bad. For each item, the degree of incongruency was indicated by the divergence between the subject's two ratings. For example, an item responded to as Mainly Like and Mainly Bad was scored 4 points incongruency; an item responded to as Mainly Like and Mainly Good received 0 points, etc. The scoring procedure was rationally determined in advance. For each subject, a *sex incongruency score* and an *hostility incongruency score* were obtained by summing the separate incongruencies for the appropriate 20 items.

Humor measures. The humor test is more fully described elsewhere (Byrne, 1957). It consists of 64 dittoed cartoons of the kind which appear in popular magazines, so selected by five judges that there are 16 cartoons in each of the following areas: hostility, sex, ridicule, and nonsense. In the present study, the cartoons were mounted individually on 6 × 8 inch cards. The subject was asked to sort them into four equal piles of 16 "Most Funny," 16 "Next," 16 "Next," and 16 "Least Funny." For each type of cartoon, a score of 3 was assigned for those placed in the most funny category, 2 for those in the next category, 1 for the next, and 0 for those rated least funny. This procedure yielded hostile, sexual, ridicule, and nonsense *humor preference scores.* The first two of these were of major interest because they corresponded directly to the incongruency measures; but, as will be seen presently, the ridicule humor preference measure was also pertinent.

Later in the testing session, the subject was given typed definitions of what was meant by sexual cartoons, hostile cartoons, cartoons which make fun of people (ridicule), and nonsense cartoons.[4] He was asked to sort all the cartoons into these four categories, putting as many as he wished into each group. From these sortings, two *humor interpretation scores* were obtained: the number of cartoons *correctly* assigned, i.e., the number of hostile cartoons categorized as hostile by the subject, the number of sex cartoons classified as sex, etc., and the *total* number of cartoons assigned to each drive area.

Procedure. Subjects were tested individually or in groups of two to four. The order of the tasks was as follows: incongruency (like-dislike), cartoons (humor preference), a sentence completion test (unrelated to the hypotheses reported here), incongruency (good-bad), and cartoons (interpretation).

TABLE 1

PRODUCT-MOMENT CORRELATIONS BETWEEN INCONGRUENCY SCORES AND HUMOR INTERPRETATION SCORES (N = 33)

Humor Identification	Hostility Incongruency	Sex Incongruency
Correct Interpretations		
1) No. hostile cartoons called hostile	−.41	−.15
2) No. sex cartoons called sex	−.10	−.32
3) No. ridicule cartoons called ridicule	.41	.27
Total Interpretations		
4) Total called hostile	−.37	−.02
5) Total called sex	−.05	−.22
6) Total called ridicule	.53	.43

Note.—One-tailed tests of significance are appropriate for correlations in Lines 1, 2, 4, and 5, for which directional predictions were made. For these a value of .296 is required for significance at the .05 level, and .409 at the .01 level. The two-tailed test is appropriate for Lines 3 and 6, where a value of .349 is required for the .05 level, and .449 for the .01 level.

RESULTS AND DISCUSSION

Humor preferences. No significant relationships, curvilinear or linear, were found between incongruency and humor preferences. Scatterplots indicated clearly that the predicted curvilinear relationships did not exist, and this conclusion was supported by a test for linearity of regression (Guilford, 1956, p. 294).

These results do not support the proposition that conflict is an essential variable in the prediction of humor appreciation. It is possible that the negative finding can be attributed to one or more aspects of the research design or the measuring instruments used. A more convincing possibility, at the moment, is that humor appreciation reflects

was followed because there is some reason to believe that persons with high F scores tend to answer personality items (such as the incongruency test) in an invalid manner. It is possible that this restriction of the sample limits the generalizability of our findings; L. Ullmann, e.g., has found, in unpublished research, that persons with low F tend to be Internalizers rather than Externalizers.

[4] The items and instructions used in the incongruency tests, and the definitions presented to the subject for the different types of cartoons have been deposited with the American Documentation Institute. Order Document No. 6693 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

attitudes, beliefs, and other habits, but not inner conflict. Further research may help to clarify this issue.

Humor interpretations. As predicted, the subjects appeared to interpret cartoon content in such a manner as to avoid percepts in areas of incongruency. Thus, the correlation between hostile incongruency and number of hostile cartoons interpreted as hostile is −.41, and the analogous correlation for sex is −.32. That is, there was a tendency for subjects having marked incongruency in hostility to avoid seeing—or at least reporting—hostile content in those cartoons which actually represented hostility. The same was true, though to a lesser degree, for the sex area. In order to hold that the tendency to avoid seeing cartoons as hostile is actually due to incongruency in the area of hostility, however, it is necessary to have some evidence that this tendency is not also related to sexual incongruency; an analogous statement can be made for the sex area. In other words, in order to support the thesis that the subjects tended to avoid percepts in a given area because of incongruency in that area, it is necessary not only to show a significant negative relationship between the number of cartoons selected and incongruency in the area, but also it is necessary that there be a low and insignificant relationship between the number of cartoons selected in that area and incongruency in other areas. As indicated in Table 1, the results tend to conform with these requirements. Thus, the correlation between number of hostile cartoons called hostile and sexual incongruency is −.15, and the correlation between number of sex cartoons called sexual and hostile incongruency is −.10.

The implication that the subjects tended to avoid percepts in areas of incongruency is further supported by correlations between incongruency in given areas and *total* number of cartoons selected for those areas. Thus, the correlation between hostile incongruency and total number of cartoons termed hostile is −.37, and the analogous correlation for sex is −.22. Further, the correlations between total number of cartoons selected in one area and incongruency in the other area are both low (−.02 and −.05), further supporting the idea of the specificity of incongruency.

Though no predictions were made concerning the relationships between the humor interpretation variables and ridicule incongruency, pertinent statistics were computed and are presented in Table 1. It may be noted that both measures of humor interpretation for the ridicule cartoons are correlated *positively* with both hostility and sex incongruency, with the relationships considerably higher for hostility. In order to interpret these findings, it is necessary to note that the ridicule cartoons can be considered as representative of indirect, disguised aggression, in distinction to the hostility cartoons, which represent direct, open aggression (Byrne, 1957). Thus, the results suggest that subjects with marked incongruency concerning hostility tend to interpret cartoons in such a manner as to obtain percepts of indirect aggression.[5] Possibly such percepts tend to satisfy to some degree the aggressive impulses without at the same time being incongruent with feelings against aggression. The positive correlations between the humor interpretation scores for the ridicule cartoons and sexual incongruency suggest that the perception of indirect aggression serves to some degree to relieve sexual incongruency also.

SUMMARY

The present study was undertaken in order to test two hypotheses about the relationships between incongruency and responses to humor. Thirty-three male neuropsychiatric patients were given an incongruency test, a humor appreciation test, and a humor interpretation test. Each contained hostile and sexual items. The results suggest that (*a*) degree of appreciation of cartoons representing a given drive area was not related to incongruency in that drive area, and (*b*) subjects tend to interpret the cartoons in such a manner as to avoid percepts in areas of marked incongruency.

REFERENCES

BYRNE, D. The relationship between humor and the expression of hostility. *J. abnorm. soc. Psychol.*, 1956, **53**, 84–89.

BYRNE, D. Response to humor as a function of drive arousal and psychological defenses. Unpublished doctoral dissertation, Stanford University, 1957.

GHOSH, R. An experimental study of humour. *Brit. J. educ. Psychol.*, 1939, **9**, 98–99.

GRZIWOK, R., & SCODEL, A. Some psychological correlates of humor preferences. *J. consult. Psychol.*, 1956, **20**, 42.

GUILFORD, J. *Fundamental statistics in psychology and education.* (3rd ed.) New York: McGraw-Hill, 1956.

LEVINE, J. Responses to humor. *Scient. Amer.*, 1956, **194**(2), 31–35.

[5] The magnitude of the correlations between the humor interpretation variables and ridicule incongruency (.41 and .53) is not due simply to a tendency of persons having considerable hostility incongruency to classify the hostile cartoons as ridicule cartoons as a way of avoiding seeing them as representing hostility. This is shown by the fact that if one eliminates the hostile cartoon from consideration completely and considers only the remaining 48 cartoons, the correlations between the two humor interpretation variables and ridicule incongruency are still sizable (.43 and .51).

McREYNOLDS, P. A restricted conceptualization of human anxiety and motivation. *Psychol. Rep.*, 1956, **2**, 293–312.

McREYNOLDS, P. Anxiety as related to incongruencies between values and feelings. *Psychol. Rec.*, 1958, **8**, 57–66.

MURRAY, H. A., JR. The psychology of humor: 2. Mirth responses to disparagement jokes as a manifestation of an aggressive disposition. *J. abnorm. soc. Psychol.*, 1934, **29**, 66–81.

REDLICH, F. C., LEVINE, J., & SOHLER, T. P. A mirth response test: Preliminary report on a psychodiagnostic technique utilizing dynamics of humor. *Amer. J. Orthopsychiat.*, 1951, **21**, 717–734.

SENF, RITA, HUSTON, P. E., & COHEN, B. D. The use of comic cartoons for the study of social comprehension in schizophrenia. *Amer. J. Psychiat.*, 1956, **113**, 45–51.

(Received November 12, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 439–441

PERCEIVED SOURCES OF SOCIAL POWER[1]

SIDNEY ROSEN
Marquette University

GEORGE LEVINGER
Bryn Mawr College

AND RONALD LIPPITT
University of Michigan

While reconciliation between the older trait theories and the more recent situational theories of leadership is not yet complete, there is growing recognition (cf. Carter, 1953; Lippitt, Polansky, & Rosen, 1952; Rosen, 1959) that there are properties of people, valued by others, which are in fact "cross-situational" sources of power. That is to say, persons derive power from the possession of attributes that have utility for others in many different kinds of situations.

For example, helpfulness in a person should be valued by others in a large variety of situations. Such an attribute, whether expressed at work, at play, or elsewhere, should lend its possessor esteem in relation to others. Ethical systems abound with descriptions of such properties, and encourage the development and manifestation of certain properties over others; yet systematic empirical knowledge regarding the identity of such properties is meager. Nor is there much scientific knowledge concerning the extent to which the layman perceives their existence.

Granted that the objective sources of social power need not, in principle, bear a one-to-one relationship to the layman's corresponding phenomenological account; it is still appropriate in the earlier stages of theory development to consult the layman, for at least three reasons: First, ample evidence (cf. Lippitt, Polansky, & Rosen, 1952) exists indicating that perceptions of social power and its sources applied to actual group situations have appreciable veridicality. Second, there is considerable evidence of a relationship between perceptions and actions, vis-à-vis social power and resources. Third, there is reason to believe that in this value-charged area the layman sometimes reveals more by his omissions than by his reporting.

The present note is an attempt to establish the identity of some cross-situational sources of power as they are perceived by persons of differing backgrounds. It is first necessary to amplify the meaning of social power and its sources (Gold, 1958; Levinger, 1959; Lippitt, Polansky, & Rosen, 1952) as we understand them. *Social power* is defined as the potential ability to influence some other person(s). We conceive of such power as stemming from the possession of *resources*, i.e., of properties that have utility for the other person or persons. In principle, a resource may be a trait, a material possession, a fund of knowledge, or even "reflected glory." The definition of resource is not restricted to properties having *positive* utility, but this distinction will not be considered here. Moreover, this paper deals with *perceived* social power and *perceived* resources.

Method

A number of boys of grade-school age were first approached in a school setting and asked, by means of open-ended interviews, to name reasons why a person could get other persons to do things. Their responses led to construction of a provisional list of 25 statements dealing with sources of power. This list was then submitted to 23 high-school age boys attending a year-round camp, with instructions to check the items that seemed most relevant and to add any that might have been overlooked. The six most frequently checked items were then administered, in the course of individual interviews, to 59 other subjects ages 12–14, in a summer camp for underprivileged and somewhat emotionally disturbed boys. The instructions were as follows:

> As you know, some guys really know how to get other guys to do things. In order to get the other guys to listen to them and do what they want, they've got to know how to go about it. Now I'd like to show you some cards with different things written on them. I want you to tell me which of these things is most important for a guy to be in getting others to do what he wants.... Which is the next most important thing....

The six items were: a guy who is fun to be with, a guy who is big and strong, a guy who knows how to do lots of things, a guy who plays fair, a guy who isn't afraid of anybody else, a guy who is always willing to help out another guy. In order to facilitate discussion, these items will be designated, respectively, by the following terms: Sociability, Physical Strength, Expertness, Fairness, Fearlessness, and Helpfulness.

[1] This study was conducted as part of a larger research study under grant (M-450) from the United States National Institute of Mental Health. Principal investigator was Ronald Lippitt, Research Center for Group Dynamics, University of Michigan.

A few months later, the same six items were presented in questionnaire form to 14 female school teachers enrolled in a graduate extension course. The instructions were similar, although the wording was less colloquial.

RESULTS AND DISCUSSION

Before proceeding to consider the actual order in which the six items were ranked, let us first examine the degree of agreement among the different judges as to the relative importance of the six resource areas for getting other persons to do things. We computed Kendall's coefficient of concordance, *w*, for the two sets of rankings (Kendall, 1948). The two coefficients showed that considerable agreement existed both among the 59 campers and among the 14 school teachers. These coefficients were .60 and .56, respectively, each being significant at beyond the .001 level.

The mean ranking given by the male adolescents was in the following decreasing order of importance: Helpfulness, Fairness, Sociability, Expertness, Fearlessness, and Physical Strength. From the viewpoint of the adult sample, the mean rank order of these items was the *same*, with the minor exception of Fairness being placed above Helpfulness. (Essentially the same results were obtained subsequently with a sample of 22 undergraduates.)

The question arises as to whether the differential importance attributed to these resource areas does not reveal that some of them are more cross-situational than others. For example, the cluster represented by Fearlessness and Physical Strength would appear to have more circumscribed utility (at least in Western cultures) than the other four resource areas.

It is reasonable to suppose, too, that the utility of Fearlessness and Physical Strength may be more limited in at least two ways: first, with respect to the *variety* of possible situations for which the properties might be relevant; and second, with respect to the *frequency* with which any given situations are likely to recur. Fearlessness and Physical Strength presumably become relevant in situations of stress or emergency. Resembling Weber's "charisma," these properties lend their owner authority under extraordinary conditions (Weber, 1947).

In order to test the inference that Fearlessness and Physical Strength were of less cross-situational utility than the other four resources, the sample of teachers was asked to rank the relative importance of the six resource areas with respect to each of six hypothetical situations: (*a*) running a successful party for club members, (*b*) conducting a successful meeting of co-workers, (*c*) raising children properly, (*d*) organizing a successful fishing trip, (*e*) dissuading an angry crowd of men, and (*f*) the successful training of commandos. It was felt that the customary ranking assigned to the six resource areas with reference to the "general" situation would be seen as least applicable in situations *e* and *f*.

It was first established that there was consensus within each situation in ranking the relative importance of the six resource areas with regard to that situation (the six coefficients of concordance were all significant at beyond the .01 level). Spearman rho's were then computed to compare the mean ranking for the *general* situation with the mean rankings of the resource areas for each of the *specific* situations. The correlations between the general ranking and those given for situations *a* through *f*, respectively, were: .60, .83, .94, .71, .09, and −.14. These correlations demonstrate that the customary ranking given to the six resource areas was applicable only to the first four situations.

More specifically, Fearlessness was ranked as the most important source of power for both the commando training and the crowd dissuading situations, while Physical Strength was ranked 3.5 and 4, respectively, for those two situations. These shifts indicated, as expected, that the relevance of Fearlessness and Physical Strength is increased in situations of stress.

It is not immediately clear why the subjects did not assign a ranking of even greater importance to Physical Strength with regard to the two stressful situations. One possibility is that the latter two situations, as depicted, were not sufficiently discriminable from the remaining situations to provide a more clear-cut test of the inference in question. Another possibility is that these stress-type situations were not appropriately geared to the characteristics of the present sample. Perhaps, if the life situations of these subjects had been closer in time (or had at least entailed exposure) to combat conditions, the subjects would have shown a greater readiness to acknowledge that Physical Strength could be a source of power (Berkum & Meeland, 1958). Still, another possibility is that strong pressures exist in our culture against acknowledging that any property capable of serving as a negative resource could be a source of power.

It should be noted, however, that consensus among these subjects in ranking Physical Strength as least important in the *general* situation was overwhelming (12 out of 14 subjects ranked it sixth), and that this degree of consensus exceeded by far the degree of consensus obtained for any of the other five resource areas.

Persons of different backgrounds thus tend to

agree about the relative importance of various sources of social power. In view of the fact that these results held up across two such widely differing samples, it is questionable whether the findings can be dismissed on the grounds of either sampling unreliability or sampling bias. One way of explaining the rank order reported here is that the highest ranked resources seem to have the greatest cross-situational utility. This interpretation leads us to propose two general hypotheses for future testing: the larger the *number* of situations for which a given resource is appropriate, the greater the perceived importance of the resource as a source of social power; the more recurrent the *kind* of situation for which a given resource is appropriate, the greater the perceived importance of the resource as a source of social power.

REFERENCES

BERKUM, M., & MEELAND, T. Sociometric effects of race and of combat performance. *Sociometry*, 1958, **21**, 145–149.

CARTER, L. F. Leadership and small-group behavior. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 257–284.

GOLD, M. Power in the classroom. *Sociometry*, 1958, **21**, 50–60.

KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.

LEVINGER, G. The development of perceptions and behavior in newly formed social power relationships. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor: Research Center for Group Dynamics, Institute for Social Research, Univer. Michigan, 1959.

LIPPITT, R., POLANSKY, N., & ROSEN, S. Dynamics of power. *Hum. Relat.*, 1952, **5**, 36–64.

ROSEN, S. Effects of adjustment on the perception and exertion of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor: Research Center for Group Dynamics, Institute for Social Research, Univer. Michigan, 1959.

WEBER, M. *The theory of social and economic organization*. New York: Oxford Univer. Press, 1947.

(Received December 24, 1959)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 442-445

INTERPERSONAL PERCEPTION SCORES AND THEIR COMPONENTS AS PREDICTORS OF PERSONAL ADJUSTMENT[1]

ALAN R. BASS AND FRED E. FIEDLER[2]
University of Illinois

CRONBACH (1958) has recommended that complex "interpersonal perception" scores, such as measures of profile similarity, be analyzed into simpler components. This type of analysis would presumably lead to a more parsimonious interpretation of the results, and indicate the extent to which the "dyadic" components of these measures contribute useful information. By "dyadic" components we shall here refer to scores that involve the comparison of two profiles. "Assumed Similarity" (AS) scores, which have been effective predictors in previous studies of leadership (Fiedler, 1958), quasi-therapeutic relations (Fiedler, Hutchins, & Dodge, 1959), and social relations (Fiedler, Warrington, & Blaisdell, 1952) are of this type. These AS scores are obtained by comparing a person's description of two others (or himself and another) by means of the *D* score (Cronbach & Gleser, 1953; Osgood & Suci, 1952).

This paper briefly summarizes a reanalysis of certain data from a study relating AS to adjustment (Bass & Fiedler, 1959),[3] to determine empirically the extent to which dyadic or pattern scores contribute to prediction and interpretation of results.

PROCEDURE

The subjects were 200 members of small antiaircraft artillery units.[4] Each subject described himself, his "ideal self," his "confidant" with whom he would discuss his personal problems, and four members of his work group. Descriptions were made on scale sheets resembling the semantic differential. Each of these scale sheets contained 20 pairs of bipolar adjectives of an evaluative nature, such as:

Friendly __:__:__:__:__:__:__ Unfriendly

Cooperative __:__:__:__:__:__:__ Uncooperative

From these scale sheets two types of AS scores could be obtained. The first of these involves the similarity a man sees between himself and others in his group—i.e., AS *by* the subject *to* others. The main AS score of this type is the $\overline{\text{AS}}$g (mean AS between subject and group members). Each subject described himself and four members of his group; an average *D* score was then computed between the subject's self-description and his description of each of the four group members. Each of these *D* scores is algebraically defined as follows:

$$\text{ASg} = \sqrt{\sum_{i=1}^{n} (X_{ip} - X_{ig})^2}$$

Where ASg is the AS between subject and particular group member, X_{ip} is the score on Item *i* for the perceiver's self-description, X_{ig} is the score on Item *i* for the perceiver's description of the group member, and *n* is the number of items in the scale sheet description. In addition to $\overline{\text{AS}}$g, two other *D* scores of this first type were computed. One was ASc, the AS between the subject's self-description and his description of his confidant. The other was ASIc, the AS between the subject's ideal-self-description and his description of his confidant. These AS scores are algebraically defined in a manner similar to $\overline{\text{AS}}$g above.

The second type of AS score involves the similarity that the subject's fellow group members perceive between themselves and the subject. The main score of this type was called $\overline{\text{AS}}$s, the average AS between the group members' own self-descriptions and their descriptions of the subject. As before, each of these *D* scores is algebraically defined as follows:

$$\text{ASs} = \sqrt{\sum_{i=1}^{n} (X_{ip} - X_{is})^2}$$

where ASs is AS to the subject by a particular group member, X_{ip} is the score on Item *i* for the self-description of the perceiver (in this case, subject's fellow group member), X_{is} is the score on Item *i* for the group member's description of the subject, and *n* is again the number of items in the scale sheet description. In addition to $\overline{\text{AS}}$s, two other *D* scores of this second type were used in the present study. One of these was C's ASs, or AS between confidant's self-description and his description of the subject; and the other was C's ASIs, or AS between confidant's ideal-self-description and his description of the subject.

AS is operationally defined by *D* scores. The *D* score can be algebraically analyzed into a number of simpler components by expanding the term under the radical, converting the raw scores to deviation scores and simplifying. This analysis indicates that:

[1] This study was supported by Research Grant M-1774, National Institute of Mental Health, United States Public Health Service.

[2] We wish to express our appreciation to S. Alexander for comments on the manuscript.

[3] A mimeographed copy of the complete report may be obtained from the authors, Department of Psychology, University of Illinois, Urbana, Illinois.

[4] These data were originally obtained in a study by Fiedler, Hutchins, and Dodge (1959).

$$\frac{D^2}{n} = (M_1 - M_2)^2 + V_1 + V_2 - 2\,\mathrm{cov}_{12}$$

where M_1 is the mean of the item scores in one of the two profile descriptions, M_2 is the mean of the items in the second description; V_1 and V_2 are the variances of the item scores within the first and second descriptions, respectively; and cov_{12} is the covariance between the item scores in the two profile descriptions. The present analysis of AS scores utilized six components for each AS score: the two means, the two variances, the covariance, and the absolute difference between the means (the "elevation difference"). Four of these are "monadic" components, involving only one of the two profile descriptions of the AS score, viz., the means (the "elevation components"), and the variances of the items within each of the two descriptions. The remaining components (the covariance and the elevation difference) are the two "dyadic" components of the AS score.

Five criterion measures indicative of the subject's adjustment were obtained: a Taylor Manifest Anxiety (*MA*) scale score, a General Army Adjustment Index, a Self-Satisfaction measure (the self–ideal-self discrepancy), the subject's sociometric acceptance by the group, and the crew leader's rating of the subject's adjustment to military discipline. A previous investigation (Fiedler, Hutchins, & Dodge, 1959) had shown that individuals who perceived others in their group as similar, and those who were seen by fellow group members as similar (*a*) tended to be better adjusted than those who saw others, and were seen by others, as relatively less similar, and (*b*) tended to improve in adjustment. Although from a substantive point of view, *a* was a relatively unimportant aspect of the original study, it is methodologically simpler for purposes of this paper. Hence, the present study reanalyzed the data relevant only to *a*, using components of the several AS scores.

RESULTS AND DISCUSSION

Factor analyses of the self-description and description of confidant scales indicated that 16 of the 20 scale sheet items were highly loaded on an "evaluative," first centroid factor. The present analyses of AS score components are based on these 16 items.

Each of the various components was correlated with the adjustment criteria. Since there were six different AS scores and five different adjustment criteria, 30 different sets of relationships between AS score components and adjustment criteria were involved. In 24 of these 30 relationships, one or more of the AS score components correlated with the adjustment criteria as high as, or higher than, did the corresponding AS score. Of these components, the profile means (or "esteem" scores) appeared to be the most promising as predictors.

TABLE 1

CORRELATIONS BETWEEN DESCRIPTION MEANS OF THE AS SCORES AND THE ADJUSTMENT CRITERIA

Mean Components	Adjustment Criteria				
	MA scale	Army Adjustment	Self-Satisfaction	Disciplinary Ratings	Sociometric Status
Subjects Description of					
Self	.30**	.27**	[a]	.21**	.13
Ideal-Self	.21**	.10	[a]	.16*	.14*
Confidant	.22**	.29**	.13	.27**	.00
Group Members	.28**	.37**	.23**	.16*	.00
Confidant's Description of					
Self	−.05	.05	.09	.21*	.04
Ideal-Self	.00	.07	.05	−.10	.06
Subject	−.09	.04	.02	.32**	.35**
Group Members' Descriptions of					
Self	−.06	−.06	.06	.05	.02
Subject	.08	−.04	.02	.21**	.45**

[a] Spurious correlations, based in part on same self- or ideal-self-descriptions.

* $p < .05$.

** $p < .01$.

In 24 of the 30 comparisons, the multiple correlations between the two description means and the criterion were as high as, or higher than, the correlation between the corresponding AS score and the criterion.

More specifically, the subject's self-description means ("self-esteem") and the means of the descriptions of his group members were most important for predicting adjustment criteria of the self-report type (*MA* scale, Army Adjustment, and Self-Satisfaction). The means of the group members' descriptions of the subject were significantly related to adjustment criteria based on "ratings-by-others" (Sociometric Status, Disciplinary Ratings). The correlations between the elevation components (means) of the various AS scores and the five adjustment criteria are presented in Table 1. It appears, then, that individuals who describe both themselves *and* their fellow group members more favorably tend to appear as better adjusted on self-report measures. Subjects who are described more favorably by their fellow group members also tend to be seen as better adjusted on sociometric indices.

To what extent do the remaining, more complex, nonlinear, and dyadic components of the AS score (the variances, covariance, and elevation difference components) contribute to prediction of criterion variance beyond that already determined

TABLE 2

CORRELATIONS BETWEEN VARIANCE COMPONENTS OF THE AS SCORES AND THE ADJUSTMENT CRITERIA

Variance Components	Adjustment Criteria				
	MA scale	Army Adjustment	Self-Satisfaction	Disciplinary Ratings	Sociometric Status
Subjects Description of					
Self	−.32**	−.14*	[a]	−.12	−.13
Ideal-Self	−.17*	−.06	[a]	−.06	−.14*
Confidant	−.25**	−.06	−.34**	−.16*	−.11
Group Members	−.26**	−.12	−.36**	−.04	−.14*
Confidant's Description of					
Self	.01	.00	−.11	−.16	.00
Ideal-Self	−.02	−.03	−.06	.12	−.13
Subject	.10	.12	.04	−.09	−.27**
Group Members' Descriptions of					
Self	−.12	.08	−.08	.03	−.14
Subject	−.20**	.10	−.18*	−.18*	−.37**

[a] Spurious correlations, based in part on same self-descriptions.

* $p < .05$.

** $p < .01$.

by the profile means? Multiple regressions were computed by successively adding to the two profile means the two variances, the covariance, and the elevation difference components, in that order, determining at each step the increment in criterion variance accounted for by the addition of that component. In only 5 of the 30 cases did the variance components add significantly to the description means in predicting adjustment measures.[5] While the profile variances tended to be related to the various criteria to about the same extent as the profile means, the means and variances based on the same profile tended to be substantially intercorrelated. Therefore, the profile variances generally add little to criterion prediction beyond that accounted for by the means. Table 2 presents the correlations between the variance components of the AS scores and the various adjustment criteria.

In four additional cases the elevation difference component added significantly to criterion prediction by the means and variances. The dyadic covariance component of the AS score tended to have the lowest correlations with the adjustment criteria, as well as the lowest regression weights in predicting the adjustment measures.

These results support Cronbach's (1958) contention that analysis of the *D* score into monadic components permits simpler and more meaningful interpretations of interpersonal perception scores than is possible with *D*. Such a research strategy should tend to yield conceptually clearer interpretations of findings obtained using these measures.

It should be pointed out, however, that these conclusions are based on relationships between AS to *specific* other persons and personal adjustment criteria. Most previous work has dealt with the score ASo, the AS between the most and least preferred co-workers the subject has ever known. This score has yielded substantial correlations with group effectiveness criteria in a number of different studies (Fiedler, 1958), and the ASo score may well yield more powerful prediction than its components in relation to group effectiveness criteria (Hutchins, 1958; Peters, unpublished).

[5] These particular results may reflect the fact that the elevation scores were entered into the multiple regressions before the variance components. If this procedure were reversed and the variance components were inserted in the multiple regression formula first, followed by the elevation components, we might then have concluded that the elevation components add little to criterion variance beyond that obtained with the variances. However, the means are obviously easier to interpret than are the variances or the more complex *D* score components. Therefore, our interest in this study was in determining whether the profile means alone are generally sufficient to account for about as much criterion variance as the *D* score can account for, and whether the addition of the more complex, nonlinear, and dyadic components of the *D* score to the profile means generally tends to increase significantly prediction of these adjustment measures.

REFERENCES

BASS, A. R., & FIEDLER, F. E. Interpersonal perception scores: A comparison of *D* scores and their components. Technical Report No. 5, 1959, Group Effectiveness Research Laboratory, University of Illinois, United States Public Health Project M-1774.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: Stanford Univer. Press, 1958. Pp. 353–379.

CRONBACH, L. J., & GLESER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–473.

FIEDLER, F. E. *Leader attitudes and group effectiveness*. Urbana: Univer. Illinois Press, 1958.

FIEDLER, F. E., HUTCHINS, E. B., & DODGE, JOAN S. Quasi-therapeutic relations in small college and military groups. *Psychol. Monogr.*, 1959, **73**(3, Whole No. 473).

FIEDLER, F. E., WARRINGTON. W. G., & BLAISDELL, F. J. Unconscious attitudes as correlates of sociometric choice in a social group. *J. abnorm. soc. Psychol.*, 1952, **47**, 790–796.

HUTCHINS, E. B. Task-oriented and quasi-therapeutic role functions of the leader in small military groups. Unpublished doctoral dissertation, University of Illinois, 1958.

OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251–262.

(Received January 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 446–449

THE GENERALITY OF ACCURACY OF INTERPERSONAL PERCEPTION[1]

VICTOR B. CLINE AND JAMES M. RICHARDS, JR.
University of Utah

In a previous paper by the authors (Cline & Richards, 1960) evidence was presented indicating that accuracy of interpersonal perception is to some degree a general trait over diverse persons (about whom judgments are made) and different measuring instruments, using a procedure in which an interview of the person to be judged is presented by means of a sound color movie. These conclusions are considerably different from those reached by Gage and Cronbach (1955), Crow and Hammond (1957), and Stone and Leavitt (1954), among others who conclude that there is little or no generality in judging ability. Because of this divergence of conclusions, a most important question is how well the generality we obtained holds up in replication, or cross-validation, studies. This is especially true since the filmed interviews used in the previous research were selected on the basis of how well they discriminated good judges from poor judges, or, in other words, how well they correlated with total judging score.

This note reports the results of four such replication studies.

METHOD

The procedure used in all of these studies was for a group of judges to view a filmed interview, in which questions were asked the person in the film by a skilled interviewer about their personality, values, etc. The persons who were in the films were quite diverse with regards to age, social status, cultural background, and included both men and women.

The judges who viewed these films filled out various judging instruments dealing with the person in each film. These instruments were constructed on the basis of intensive interviews, both with the person in the film and with five of his close associates, and a battery of psychological tests. The judges saw a filmed interview, then the projector was stopped, and the judging instruments for that film filled out. Then the second film was shown, the projector stopped, and the judging instrument for that film filled out, etc. The experimental procedures and the development of the judging instruments are described in detail elsewhere (Cline & Richards, 1960).

Four different judging instruments were used in these replication studies. These were: the Behavior Postdiction test, where the judge was required to postdict the interviewee's real life behavior on a seven-item-per-film multiple-choice test; the Trait Rating test, where the judge rated the interviewee on 25 adjectives using a six-category Likert scale; the Opinion Prediction test, where the judge predicted how the interviewee had responded to 20 selected MMPI items; and the Adjective Check test where the judge had to guess which one of a pair of adjectives each interviewee had checked as being descriptive of himself for each of 20 pairs of adjectives. On each of these four instruments items were designed so that the subjects could not obtain a high score on the basis of always checking the most socially desirable alternative or through some other similar response set. Precautions were taken to choose interviewees very heterogeneous as to age, social-cultural background, sex, etc. In addition, a Total Judging score was computed by converting scores on individual instruments to standard scores and adding across instruments. Due to difficulties in scheduling and/or motivating subjects, not all of these four judging instruments were used in all of the replication studies.

RESULTS

Subjects in the first replication study consisted of 295 students of both sexes recruited from elementary psychology classes at the University of Utah in the Fall of 1958. These subjects were tested using the 10 films employed in our original study (which gave evidence for the generality of judging ability). In an effort to reduce the testing time, a shorter form of 6 films was selected from these 10 films, with selection again being made on the basis of how well the film discriminated poor judges from good judges. Intercorrelations of the accuracy scores on the different judging instruments for these subjects for both the 10-film form and the 6-film form are presented in Table 1. In this and all subsequent tables, all correlations between individual judging instruments and Total Judging are part-whole correlations corrected for the contribution of the part, through use of the usual correction formula (Gulliksen, 1950).

Subjects in the second replication study consisted of 51 students of both sexes recruited from upper division psychology classes at the University of Utah in the Spring of 1959. These subjects were tested with six films. One film, however, had been changed from the study presented in Table 1. In this case, selection was *not* made on the basis of discrimination of good judges from poor, but rather because the person in the film was *dissimilar* to the other five persons (films), both generally in personal characteristics and specifically on the Trait Rating instrument. At the same time Trait Rating was revised so that the adjectives included were those on which persons in the film

[1] This study was supported by Contract NONR 1288 (04), Group Psychology Branch, Office of Naval Research.

TABLE 1

INTERCORRELATIONS BETWEEN ACCURACY SCORES ON DIFFERENT JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS TESTED IN FALL QUARTER, 1958 (N = 295)

	10-Film Form			6-Film Form		
	Trait Rating	Adjective Check	Total Judging	Trait Rating	Adjective Check	Total Judging
Behavior Postdiction	.24	.04	.31	.36	.28	.39
Trait Rating		.31	.44		.41	.50
Adjective Check			.31			.44
		r .05 = .11		r .01 = .15		

TABLE 2

INTERCORRELATIONS BETWEEN ACCURACY SCORES ON DIFFERENT JUDGING MEASURES FOR UNIVERSITY OF UTAH STUDENTS TESTED IN SPRING QUARTER, 1959, USING REVISED SIX FILM FORM (N = 51)

	Trait Rating	Opinion Prediction	Adjective Check	Total Judging
Behavior Postdiction	.16	.29	.37	.37
Trait Rating		.31	.13	.28
Opinion Prediction			.25	.32
Adjective Check				.36
r .05 = .27		r .01 = .35		

TABLE 3

INTERCORRELATIONS BETWEEN ACCURACY SCORES ON DIFFERENT JUDGING MEASURES FOR IDAHO STATE HOSPITAL PATIENTS TESTED WITH REVISED SIX-FILM FORM (N = 32)

	Trait Rating	Adjective Check	Total Judging
Behavior Postdiction	.41	.61	.67
Trait Rating		.31	.37
Adjective Check			.57
r .05 = .35		r .01 = .45	

TABLE 4

INTERCORRELATIONS BETWEEN ACCURACY SCORES ON DIFFERENT JUDGING MEASURES FOR T GROUP MEMBERS TESTED WITH REVISED SIX FILM FORM (N = 118)

	Trait Rating	Opinion Prediction	Adjective Check	Total Judging
Behavior Postdiction	.28	.36	.36	.52
Trait Rating		.27	.03	.26
Opinion Prediction			.07	.47
Adjective Check				.35
r .05 = .17		r .01 = .23		

TABLE 5

RELIABILITIES OF JUDGING INSTRUMENTS (N = 140)

Behavior Postdiction	.37
Trait Rating	.60
Opinion Prediction	.59
Adjective Check	.61
Total Judging	.54

were relatively *dissimilar*. Both of these changes make it less likely that any generality of judging ability found using these films and this revision of Trait Rating can be attributed to real similarity of the persons judged to each other.

The correlations for these subjects using these six films are presented in Table 2.

Subjects in the third replication study consisted of 32 mental patients at the Idaho State Hospital South, Blackfoot, Idaho. These patients represented a variety of diagnostic categories, and were selected primarily on the basis of being sufficiently intelligent, motivated, and in contact with reality to complete the judging task. They were tested using the revised six-film form used in the second replication study.

The correlations for this group are presented in Table 3.

Subjects in the fourth replication study consisted of 118 members of "T groups" obtained from rather heterogeneous sources. By T groups we mean groups of people, 8–15 in number, who had participated in special training in group dynamics. Thirty-six subjects were students from T groups at the Brigham Young University, Provo, Utah; 17 subjects were adults from T groups at the Intermountain Training Laboratory at Cedar City, Utah; and 65 subjects were adults from T groups at the National Training Laboratory at Bethel,

Maine.[2] All of these subjects were tested using the revised six-film form used in the second and third replication studies.

Correlations for these groups, combined into a single group, are presented in Table 4.

The data presented so far have dealt with generality across judging instruments. Another important question is to what extent is there generality within instruments across persons, or, in other words how reliable are the various judging instruments. To put it another way, is a judge who is accurate in judging one person also accurate in judging other persons on the same instruments? In order to answer this question, reliability coefficients were computed using as subjects 149 judges from the second replication University of Utah student group, the Bethel National Training Laboratory, and Brigham Young University T groups. These subjects were tested with all four judging instruments using the revised six-film version of our judging test. These reliabilities are presented in Table 5. In Table 5 the reliabilities of Behavior Postdiction, Opinion Prediction, and Adjective Check are Kuder-Richardson Formula 21 coefficients, and the Trait Rating reliability is a Spearman-Brown coefficient based on a first three films-second three films split. Because of the computational labor involved in computing a split-half reliability for Total Judging, this reliability was estimated from the formula presented by Mosier (1943) for determining the reliability of a composite, using unit weights for the parts and disregarding the correlations between the parts.

DISCUSSION

One of the most important scientific criteria for evaluating research findings is whether it is possible to replicate the results. In practice, however, this criterion is somewhat difficult to apply to the results of the present studies, since not all of the individual correlations are statistically significant. These difficulties are compounded by the fact that there is some question as to whether Type I error (rejecting the null hypothesis when it is true) or Type II error (accepting the null hypothesis when it is false) is more serious in the reported investigations of the generality question. Nevertheless, in all of the studies reported here there seems to be a consistent pattern of low positive correlations between different judging instruments, and somewhat larger positive correlations between judging instruments and Total Judging (where the contribution of the particular instrument has been eliminated or partialed out). The average interinstrument correlation for the six-film form (used in the second, third, and fourth replication studies), computed through use of Fisher's z coefficient, is .25 and the corresponding average part-whole correlation is .41. Both of these correlations are significant at beyond the .001 level. It will also be noted that all 38 correlations reported here are positive. While this result is obviously relevant to the question of whether or not there is generality among the various judging instruments, we know of no really satisfactory way to evaluate it statistically. As a rough approximation, the binomial probability of obtaining all one sign in a random distribution of signs with 38 cases is also much less than .001.

The reliabilities obtained for the judging instruments are not extremely high. It should be noted, however, that Kuder-Richardson Formula 21 reliabilities are a lower bound estimate of the reliability (Gulliksen, 1950), and that the Trait Rating and Total Judging reliabilities obtained in this research are also lower than the maximum reliabilities for the instruments involved, since no attempt was made to weight in such a way as to maximize reliability. We therefore consider these reliabilities to be consistent with the correlations between instruments.

The relatively low reliability for Postdiction can be substantially attributed to the fact that it has only about one-third as many items (42 vs. 120) as the other instruments. The standard formula for relating test length to reliability (Gulliksen, 1950) indicates that if Postdiction were three times as long, and thus had approximately as many items as the other judging instruments, its reliability would be .51, a value more comparable to the reliabilities of the other judging instruments.

We regard these results as confirming the previous findings of generality of accuracy of interpersonal perception. The consistency of results over the four replications is stronger evidence than is a statistical test of the significance of correlation coefficients within a given replication. The absolute size of the correlations between judging instruments has decreased from that obtained in the original study (Cline & Richards, 1960), as could be expected in a cross-validation such as this. The correlations between judging instruments are, for the most part, less than the limits set by their reliabilities, a finding that tends to confirm the view that judging ability is factorially complex.

In the course of our research considerable effort was devoted to the treatment of scores on the Trait Rating in an analytic way suggested by Cronbach (1955), since our original study had suggested the importance of this type of analysis in studies of the generality of accuracy of interpersonal perception. Two separate programs have been written for the IBM 650 computor, one for

[2] Great appreciation is expressed to Jack Gibb for his assistance and cooperation in collecting this data.

obtaining the total score and the four basic components in difference score form, and one which performs the same computations but in addition breaks down two components into a correlation term and a variance term. These programs were tailored to our instrument, and were not written in the general case. Work with these programs indicated that due to some of the special characteristics of this analytic treatment of scores, and the particular form of the Trait Rating test used in these studies, not all of the components could be meaningfully interpreted. Because of these limitations, we made no further analysis of these components. These limitations do not, however, appear to prevent the total score on Trait Rating from being meaningful, and this score was retained in the reported studies. Trait Rating has subsequently been revised to eliminate these difficulties of interpretation, and we are currently completing a study of the characteristics and interrelations of the various components on a new group of subjects.

SUMMARY

Results of our previous research indicated that there is some degree of generality in accuracy of interpersonal perception or judging ability. Four replication studies of the original research were carried out, with the measures of judging ability being somewhat revised after the first of these studies. In the three subsequent studies using the revised measures, the average interinstrument correlation between different measures of judging ability was .25 and the average corrected part-whole correlation between individual judging instruments and a Total Judging score was .41. Since both of these correlations are significant at the .001 level, we conclude that these studies confirm the previous findings of a low but consistent degree of generality of judging ability. The reliabilities of the judging instruments, while not extremely high, also tend to indicate that there is some generality of judging ability.

REFERENCES

CLINE, V. B., & RICHARDS, J. M. Accuracy of interpersonal perception—A general trait? *J. abnorm. soc. Psychol.*, 1960, **60**, 1–7.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CROW, W. J., & HAMMOND, K. R. The generality of accuracy and response sets in interpersonal perception. *J. abnorm. soc. Psychol.*, 1957, **54**, 384–390.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GULLIKSEN, H. *Theory of mental tests.* New York: Wiley, 1950.

MOSIER, C. I. On the reliability of a weighted composite. *Psychometrika*, 1943, **8**, 161–168.

STONE, G. C., & LEAVITT, G. S. *Generality of accuracy in perceiving standard persons.* University of Illinois, 1954. (Mimeo)

(Received January 18, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 450–453

THE EFFECTS OF PRIOR INFORMATION ON SUSCEPTIBILITY TO AN EMOTIONAL APPEAL

PAUL C. LEWAN AND EZRA STOTLAND
Wenatchee Valley College *University of Washington*

CONSIDERABLE research has been done on the order of presentation of persuasive communications. A central problem of the research has been whether the earlier communication has any effect in reducing the effectiveness of the later communication (Hovland, 1957). In these studies, the communications have typically consisted of a list of arguments to support a given point of view of some object or issue. The present study is addressed to the corollary issue of whether a communication of neutral material about an object can influence the effectiveness of subsequent communications. In the present study, the second communication is primarily emotional in its appeal. In other words, the problem is, "Can the prior acquisition of neutral information concerning some object influence susceptibility to a subsequent appeal regarding that object?"

Studies such as Hartman's (1936) and Janis and Feshbach (1953) have shown at least partially contradictory results. Hartman found that emotional appeals are more effective, while Janis found that extremely emotional appeals are not effective. One possible explanation for the somewhat inconsistent results may lie in the lack of prior information that the subjects had about the object of the appeal. The students of the Janis and Feshbach study may have been better informed about dental hygiene than the voters in Hartman's study were about public issues. The importance of prior information is also suggested by Cantril's (1940) finding that better educated persons, who were probably better informed, were less influenced by emotionally tinged communications. However, it remains unclear as to whether the lowered susceptibility to influence is a result of a greater fund of information or of greater critical ability. The present study is aimed at showing that regardless of the subject's critical ability, his possession of more information does in fact reduce the affectiveness of an emotional appeal. The following hypothesis was tested:

Persons who have received neutral information regarding a given object lower their evaluation of the object to a lesser extent after hearing an emotional appeal against it than persons who have not received such neutral information.

METHOD

Subjects. Ninety-seven pupils in four twelfth grade social studies classes were used as subjects for this experiment. Approximately 52% of the subjects were male. The pupils in each of the four classes were seated in alphabetical order. No ability grouping by class was apparent to the experimenter; teachers described the classes in each case as "about average."[1]

Experimental conditions. The experimental groups received an information sheet on the same subject matter area as the emotional appeal; the control group received information on a nonrelevant subject. The possession of relevant prior information by the experimental subjects was the only condition in the design which distinguished the experimental groups from the control groups.

Each class was divided into an experimental and a control group of approximately equal size. Since all classes used alphabetical seating arrangements, subjects were assigned alternately to the experimental or control group by alphabetical order.

Procedure. The experiment can be divided into four major steps: the presentation of the fact sheets, administration of the first attitude scale, presentation of the emotional appeal, administration of the second attitude scale. The experiment was given four times on the same day, once in each of the four twelfth grade social studies classes.

First, the teacher in charge of each class introduced the experiment by informing the students that they were to study a little-known area of Europe in a way somewhat different from the usual classroom routine. The pupils were told that there were to be some tests administered after they had an opportunity to look at some information sheets. The teacher, with the experimenter acting as an observer, handed out the "fact sheets" to the experimental and control groups. The experimental group received only information on Andorra (described below). The control group received only information of a similar nature about Etruria. The fact sheets consisted of items of information designed to be of a "neutral nature" on a subject matter area with which the students would not ordinarily be acquainted. From the time of the handing out of the information until the end of the experiment, intercommunication between pupils was not permitted. The subjects were instructed to read the fact sheets; when the teacher had observed that all subjects had finished, the sheets were recalled.

Immediately after the reading of the fact sheets the teacher passed out scales exploring the attitude of both groups in the relevant subject matter area (the scales and their construction will be discussed more fully later). Half of each group received Scale A, the other half received Scale B on the initial attitude exploration.

[1] The authors wish to express their appreciation to the teachers of Evergreen High School, Seattle, Washington, for their invaluable cooperation.

The scales in every case were identified by number only, the person responding remaining anonymous.

After the completion and handing back of the first attitude scales, the experimenter was introduced by the teacher as a guest speaker from the University of Washington who would like to present a few remarks about Andorra. The experimenter immediately presented the emotional appeal, which consisted largely of name-calling generalities presenting an unfavorable picture of Andorra and had been edited to present a minimum of factual information.

After the appeal the speaker retired once more to the role of observer and the teacher gave out the alternate attitude scales to both groups. Those students who had been given Scale A before the emotional appeal were now given Scale B and vice versa.

A comparison of the "change scores" (the difference between scores on the first and second attitude scales) for the experimental and control groups was used to test the hypothesis.

The attitude scales. In order to detect the expected changes of attitude in both the experimental and control groups, two "Likert type" attitude scales were constructed, using items similar to the sample shown below:

If an Andorrese (a citizen of Andorra) was coming to your school and you were asked if you would be a "buddy" until this student becomes better acquainted would you:

1. approve
2. disapprove
3. strongly disapprove
4. strongly approve

The Andorrese is basically an honest, hard-working person. Do you:

1. strongly disagree
2. disagree
3. agree
4. strongly agree

Each of the possible four responses for the 22 items on each scale was assigned a value of from one to four with the most unfavorable response (towards Andorra) being assigned the lowest value, the next most unfavorable responses being assigned the second lowest value, etc. The response categories in the scales were randomized in their order for each item to minimize the effects of position responses by the subjects.

The two forms, designed to be used as alternate scales, had been previously administered to a group of 26 eleventh graders (different from the experimental group) in a social studies class after the students had read the information sheet on Andorra. This class was characterized as an average class by several teachers and was composed of an equal number of boys and girls at the time of the testing, which was 5 weeks previous to the runnning of the experiment. The students were informed by their teacher, immediately before the passing out of the fact sheets, that this was to be a study of a little-known country of Europe and that some tests were to be administered after the reading. After the reading of the fact sheet, about half the class received the scales in the order AB; the other half received scales in the order BA. After completion of the scales the experimenter answered a few general questions about the subject matter area. The students were not informed as to the intended use of the attitude scales in the later experiment.

The mean score on Scale A was 63.08, the mean score on Scale B was 62.73. A product-moment correlation of .91 (without correction for attenuation) was found to exist between scores on the two forms.

Fact sheets. The information sheet on Andorra received by the experimental groups was abstracted from *Collier's Encyclopedia* and carefully edited so that the items would be of a "neutral nature," i.e., unlikely to form marked attitudes towards or against the subject. The neutral information consisted of the following: a description of the geography, the agriculture, the industry and mining, the language and the use of the flag, the political system and the French representation in foreign affairs, the fiscal system, and the country's origin. The paragraphs on the political system and the fiscal system are cited here as examples:

> The internal affairs are handled by a General Council of Notables, four from each of the country's six parishes, elected by the heads of the landowning families. France usually represents Andorra in foreign affairs and has the right to maintain law and order in emergencies within the country.
>
> The country has a small and balanced budget and uses both French and Spanish currencies since there is no such thing as Andorran coinage or money. Postal service is free within the country, being financed by the sale of Andorran stamps to stamp collectors. The Andorrese have almost no taxes but pay a small tribute every two years to the Bishop of Urgel (Spain) and the President of France, a traditional tribute which has been paid for almost 700 years.

The control group received information of a similar "neutral nature," on Etruria which was condensed from another work in the same source. The two fact sheets can be read in approximately the same length of time (about 8 minutes).

Emotional appeal. The second "communication," the emotional appeal, was presented orally by the experimenter. It consisted largely of name-calling generalities and was designed to present a minimum of factual information while presenting an unfavorable opinion of the relevant subject matter area (Andorra). The emotional appeal consisted of an attack on the Andorrese people as being ignorant, amoral, unworthy of membership in the United Nations or of American economic aid, undemocratic, weak, puny, potential dupes of the Communists in the United Nations, and dangerous to world peace. A typical section follows:

> We cannot, today, gamble nor compromise with our principles. The so-called "citizens" of Andorra are known throughout Europe as backward, ignorant, unprincipled, provincial, undemocratic people. They are notorious for being poverty-stricken ne'er-do-wells, uncaring of what happens outside their principality, unchanging, feudal, cold and remote to strangers. The Arkansas hillbilly would be a sophisticated "man of the world" compared to the average Andorrese. It is revealing that the national economy is based upon immorality, is based upon smuggling, blackmarket, and theft to a large extent. In fact, if

you were to ask a well-informed person what is the one thing for which Andorra is noted, the answer would probably be "the absence of morality." In these times, when the world appears to be teetering on the brink of disaster, can we afford to give power to people of this sort? The answer must be no, we cannot and must not give to this degenerate country and its immoral people the power to manipulate our lives and destroy our freedom.

RESULTS

The hypothesis of the study is based on the assumption that the information provided to the subjects prior to the initial appeal was neutral with respect to the object of the appeal, Andorra. This assumption does not, however, require that every bit of information given be neutral; rather the total effect of the information should not cause the recipients as a group to be more or less favorable to Andorra. The assumption can be tested by comparing the experimental and control groups in their attitudes toward Andorra as measured immediately after they received information about Andorra and Etruria and before hearing the emotional appeal. The mean attitude scale score was 60.54 ($N = 47$) for the experimental groups and 59.22 ($N = 50$) for the central groups, the difference not being significant by a t test. (The higher figure indicates more favorable attitudes.) Thus, the assumption about the neutrality of the information is supported.

The hypothesis was tested by first dividing the sample into three equal groups according to degree of favorableness of attitude toward Andorra prior to the emotional appeal, to control for the differential possibility of change from different positions of favorableness of attitude, and to check on any interaction effects between the experimental treatments and original degree of favorableness. The mean change scores were then computed for each of the three groups in each of the experimental conditions. These data are presented in Table 1. They show that for each degree of original favorableness, the subjects in the experimental condition changed less after the negative emotional appeal than subjects in the control condition toward being more unfavorable to Andorra. The analysis of variance shown in Table 2 indicates that this difference is significant at the .05 level. The analysis of variance also shows no significant interaction, indicating that the experimental effect was the same for subjects at all levels of original attitude. The hypothesis is therefore supported. The acquisition of neutral information about Andorra did serve as a deterrent to susceptibility to an emotional appeal against Andorra.

TABLE 1

MEAN BEFORE-AFTER CHANGES IN ATTITUDES[a] TO ANDORRA

($N = 10$ in each cell[b])

Condition	Favorableness or Attitude Prior to Appeal		
	High	Medium	Low
Experimental	−12.7	−9.5	−8.8
Control	−17.5	−18.1	−12.5

[a] Negative changes mean changes to more unfavorable attitudes.

[b] Some Ns were dropped to equalize the Ns in the cells. The Ns to be dropped were those closest to the mean in each cell before equalization, thereby increasing the variance.

TABLE 2

ANALYSIS OF VARIANCE OF CHANGES OF ATTITUDE

Source of Variation	SS	df	MS	F
Experimental Condition	478.3	1	478.3	5.49*
Original Attitude Position	209.4	2	104.7	NS
Interaction	66.1	2	33.1	NS
Within Cells	4793.1	54	88.8	NS

* $p < .05$.

DISCUSSION

One interpretation of these results would hold that the prior information creates a cognitive structure with regard to the object of the attitude, which then serves as a source of attitudinal constancy, since any change in the evaluation of the object tends to require support from the cognitive structure regarding the object. To the extent that such a cognitive structure contains elements inconsistent with the changed evaluation of the object, the person should resist changing his evaluation of the object of the attitude. This interpretation is consistent with the approach of Krech and Crutchfield (1948).

The results of this study concerned resistance to changes in a negative direction. The obvious question arises as to whether such information would also act as a deterrent to changes in a positive direction. If the process underlying the presently observed effect is simply, as just suggested, that information provides an anchor or a basis of constancy of attitude, then the information should prevent positive changes as well. However, other processes may underlie the effect. For example, as a result of their familiarity with Andorra, the subjects may have identified with it, and, thus, become resistant to any negative appeal about Andorra. If this interpretation is valid, it might be expected that the better informed subjects would then be more susceptible to a positive

emotional appeal than the less well informed ones. The importance of the factor of familiarity with, and consequent identification with, the object of the attitude can be tested by a study in which familiarity is systematically varied without varying the amount of information given about the object. In any case, it is possible that the present results can be generalized only to resistance to negative appeals and not to resistance to positive appeals.

Another possible limitation to the generalizability of the findings derives from the fact that it is impossible to make an emotional appeal without at the same time communicating some "information," even if it is minimal and/or false. In some cases, this "information" may be contrary to the originally imparted, neutral information. In this case, the results would be the same as in the present study, the original information presumably having the greater effect, as in Luchin's (1957) study. In other cases, the emotionally communicated "information" may be consistent with the prior, neutral information, or may give the neutral information an additional context in which the prior information is no longer neutral but supports an extreme pro or con attitude. In these cases, the outcome would probably be the opposite of the present study, the prior information increasing the likelihood of later susceptibility to an emotional appeal.

SUMMARY

The hypothesis of this study was that persons who have received neutral information with respect to an object lower their evaluation of the object to a lesser extent after hearing an emotional appeal against it than persons who have not received such neutral information. The hypothesis was tested in an experiment done in high school classes, using an object with which the subjects had little or no prior acquaintance. The results confirmed the hypothesis.

REFERENCES

CANTRIL, H. *The invasion from Mars.* Princeton: Princeton Univer. Press, 1940.

HARTMAN, G. W. A field experiment on the comparative effectiveness of 'emotional' and 'rational' political leaflets in determining election results. *J. abnorm. soc. Psychol.*, 1936, **31**, 99–114.

HOVLAND, C. (Ed.) *The order of presentation in persuasion.* New Haven: Yale Univer. Press, 1957.

JANIS, I., & FESHBACH, S. Effects of fear-arousing communications. *J. abnorm. soc. Psychol.*, 1953, **48**, 78–92.

KRECH D., & CRUTCHFIELD R. *Theory and problems of social psychology.* New York: McGraw-Hill, 1948.

LUCHINS, A. S. Primacy—recency in impression formation. In C. Hovland (Ed.), *The order of presentation in persuasion.* New Haven: Yale Univer. Press, 1957.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 454-457

ANTI-SEMITISM AND THE EXPRESSION OF FANTASY AGGRESSION[1]

DONALD WEATHERLEY[2]
University of Colorado

ACCORDING to the scapegoat theory of prejudice, the manner in which a person copes with his aggressive urges has an important bearing on the degree to which he exhibits prejudicial reactions toward members of minority groups. The prejudiced person is assumed to be apprehensive about directly aggressing toward the real instigators of his hostility and as a consequence tends to express aggression in indirect forms and/or against substitute objects—i.e., he displaces aggression. This tendency to displace aggression is considered an important factor in producing and maintaining ethnic prejudice (e.g., Dollard, Doob, Miller, Mowrer, & Sears, 1939).

If there is in fact a functional relationship between a tendency to displace aggression and degree of prejudice we should expect to find that intolerant persons have a greater tendency than tolerant persons to displace aggression. Studies bearing directly on this proposition (Berkowitz, 1959; Lindzey, 1950) have produced inconclusive results. The pertinence of these studies to the scapegoat theory involves the assumption that prejudice is but one aspect of a more general tendency to displace aggression on the part of intolerant individuals. As Lindzey (1950) first suggested, however, the prejudiced person may be one who has acquired a specific tendency to displace aggression *selectively* toward members of certain minority groups. Thus, for example, the highly anti-Semitic individual may have a strong tendency to displace aggression onto Jews, but may not necessarily show other evidence of a strong tendency to displace aggression. The present study was designed to test this hypothesis.

Subjects differing in degree of anti-Semitism were compared with respect to their readiness to express experimentally aroused aggression in subsequent fantasy productions. By a special manipulation of the instructions for the fantasy measure, it was possible to obtain measures of aggression expressed toward fantasy characters with Jewish or non-Jewish names. As a test of the hypothesis that High A-S subjects have a greater specific tendency than Low A-S subjects to displace aggression selectively onto Jews, it was predicted that as a function of aggression arousal: High A-S subjects should express more fantasy aggression toward Jewish-named characters than Low A-S subjects; and High A-S subjects, as compared with Low A-S subjects, should direct more aggression toward fantasy characters with Jewish names than toward characters with non-Jewish names. The second prediction, it can be seen, refers to an interaction between the variables of A-S level and character name.

METHOD

Subjects and Procedure

The data were obtained from 100 male, non-Jewish college students, selected on the basis of extremely high and extremely low scores on the Levinson Anti-Semitism (A-S) Scale. Twenty-five High A-S subjects and 25 Low A-S subjects were subjected to an aggression arousing situation, and a specially designed fantasy test was then administered. Control subjects, individually matched with the angered subjects on the basis of A-S scale scores, took the fantasy test without prior aggression arousal.

The data were collected in small group sessions. In Arousal group sessions aggression arousal was accomplished by making highly insulting and depreciating comments to the subjects in the course of administering a brief paper-and-pencil task to them. The insulting was done by the author. Immediately after this procedure a different experimenter administered the fantasy test. The situation was arranged so that the subjects would not perceive a relationship between the aggression arousing situation and the subsequent fantasy test. Following the fantasy test and before the subjects were told of the deception employed, they were asked to complete an open-ended questionnaire that gave them an opportunity to express hostility toward the author.

The same procedure was followed for the control subjects, with the exception that the paper-and-pencil task was administered in a friendly, nonprovoking manner.

A more complete description of this procedure has been reported elsewhere (Weatherley, 1956).

Fantasy Test

Eight original pencil sketches, each depicting four male figures involved in an ambiguous situation, were successively projected on a screen for 20 seconds in constant order in all sessions.[3] The subjects were given 4 minutes to write a story about each picture.

[1] This paper is based on a doctoral dissertation done at Stanford University. The work was done while the author was on active duty in the Army Medical Service Corps, enrolled in the Army Senior Psychology Training Program.

[2] The author gratefully acknowledges the aid given by J. K. Adams, Robert Sears, and C. L. Winder.

[3] Photographs of these pictures have been deposited with the American Documentation Institute. Order Document No. 6741 from ADI Auxiliary Publications

In addition to the usual group TAT instruction (McClelland, Atkinson, Clark, & Lowell, 1953) the subjects were given the age, name, and occupation of a fictional character at the time each picture was shown. They were instructed to tell what was happening to that character in their story about the picture. Four characters were named and described to the subjects. Thus each character was presented twice in the series of pictures.

Two of the characters were given Jewish sounding names: Sam Goldblatt and Herb Rosen. The other two characters were called Ken Taylor and Jim Brooks. These were considered non-Jewish names. The choice of names was based on empirical results reported elsewhere (Weatherley, 1956). Each character was not only named, but described to the subjects as either a 28-year-old college graduate student or a 16-year-old high school student. Each of these two descriptions was combined equally often with Jewish and non-Jewish names.

Analysis of Fantasy Data

For each subject, a count was made of the number of aggressive acts directed toward each of the four characters which the subjects had been told to include in their stories. From this count, two total scores were derived for each subject. These scores corresponded to the total number of aggressive acts directed toward Jewish and non-Jewish characters, respectively.[4] The data were then reduced to difference scores by subtracting the two scores of each control subject from the comparable scores obtained by his matched partner in the Arousal group. Since the Arousal and Control groups differed only in that the former was subjected to an aggression arousing experience prior to taking the fantasy test, these A-C difference scores reflected the net effect of aggression arousal on fantasy aggression.

RESULTS

Effectiveness of Aggression Arousal

The questionnaire administered immediately after the fantasy test and before the subjects were told of the deception provided one basis for comparing the reactions of the angered subjects with that of control subjects. As evaluated by three raters, 65% of the Arousal group subjects were judged to have expressed hostility toward the experimenter who had insulted them, while none of the subjects in the Control group did so. (There was not a significant difference within the Arousal group between High and Low A-S subjects on this measure.) In addition to this questionnaire, at the end of the Arousal group sessions the subjects were asked to describe in writing how they had felt when insulted earlier in the session. All but 6 of the 50 Arousal group subjects admitted that they had felt angered by the comments of the experimenter. In addition, all of these subjects stated that they did not perceive a relationship between the aggression arousing situation and the subsequent fantasy test. Thus both in the report of their subjective feelings and in an objective measure of hostility expressed, the Arousal group subjects gave convincing evidence that aggression arousal had been effective.

TABLE 1

A-C MEAN DIFFERENCES IN NUMBER OF AGGRESSIVE ACTS TOWARD CHARACTERS WITH JEWISH AND NON-JEWISH NAMES

Object of Aggression	Pairs of High A-S Subjects ($N = 25$)	Pairs of Low A-S Subjects ($N = 25$)	t between Means
Jewish Characters	.28	−2.04	3.05***
Non-Jewish Characters	−.56	−.68	.19
t between Means	1.81*	2.26**	

* $p < .05$, using a one-tailed test.
** $p < .05$.
*** $p < .01$.

Effect of Aggression Arousal upon Fantasy Aggression

The net effect of aggression arousal upon fantasy aggression is reflected in the A-C mean differences reported in Table 1. It can be seen that for High A-S subjects, aggression arousal produced a slight increase in fantasy aggression toward Jewish-named characters and a slight decrease in aggression toward characters with non-Jewish names. Neither of these A-C means differed significantly from zero. They did, however, differ significantly ($p < .05$) from one another when a one-tailed test was used.

For Low A-S subjects aggression arousal resulted in a very significant decrease in fantasy aggression directed toward Jewish-named characters ($t = 3.85$; $p < .001$) but only a slight, nonsignificant decrease in aggression directed toward non-Jewish characters. These two A-C means also differed significantly ($p < .05$) from one another.

As the findings reported above suggest, aggression arousal did have a differential effect upon the fantasy aggression of High and Low A-S subjects. It resulted in a significant difference ($p < .01$) between High and Low A-S subjects in the amount of aggression directed toward fantasy characters with Jewish names. An analysis of variance of the A-C difference scores also revealed a significant interaction ($F = 8.35$; $p < .01$) between the variables of anti-Semitism and "Jewishness" of the characters selected as targets of fantasy aggres-

Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.25 for microfilm or $1.25 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

sion. These obtained differences between High and Low A-S subjects were in accord with the predictions made.[4]

DISCUSSION

Despite evidence that the aggression arousing situation was successful in making the subjects angry, it did not result in a significant increase in fantasy aggression—none of the A-C means was significantly positive. Displacement of aggression in fantasy form was thus not formally demonstrated to have occurred. In general the reverse was true, especially in the case of the Low A-S subjects. Lindzey (1950), however, was able to demonstrate an increase in the fantasy aggression of both tolerant and intolerant individuals following a frustating situation. This discrepancy between Lindzey's results and the present findings suggests that situational factors may play an important role in determining the net effect of aggression arousal upon subsequent, indirect expressions of aggression.

In contrast to the frustration procedure used by Lindzey, the present study used a technique that evoked anger directed toward an easily identified instigator, implicitly involved strong sanctions against counteraggression, and evoked anger in subjects to an extent that made it relatively easy for them to recognize their own hostile feelings. Factors such as these may well have led to inhibiting influences upon subsequent fantasy aggression, thus reducing the average level of fantasy aggression in the angered subjects. In any event, the results dramatically demonstrated that aggression arousal did produce differences in reaction to the subsequent fantasy test.

In terms of the hypothesis under study, an important finding was that aggression arousal resulted in the High A-S subjects directing significantly more aggression toward Jewish characters than they did toward non-Jewish characters. This finding suggests that highly anti-Semitic persons do tend to select Jews over non-Jews as targets for aroused aggression. Furthermore, as predicted, it was shown that High and Low A-S subjects differed reliably in this respect. Also as predicted, High A-S subjects directed significantly more aggression toward the Jewish-named characters than did the Low A-S subjects as a function of aggression arousal. These results then support the hypothesis that persons high in anti-Semitism have a greater specific tendency than persons low in anti-Semitism to displace aggression selectively onto Jews.

Another aspect of the data serves to underscore the specific nature of the differences found between prejudice groups. High and Low A-S subjects did not differ significantly as a function of aggression arousal in the amount of aggression expressed toward the non-Jewish characters. The implication is that at least in a situation where Jewish objects are available as targets for aggression, the High A-S subject is not any more likely than the Low A-S subject to direct aroused aggression toward non-Jewish substitute objects. This finding argues against an explanation of prejudice as representing merely one facet of a more general tendency to displace aggression diffusely.

So far, Low A-S subjects have been discussed only in connection with the differences found between prejudice groups. However, the reaction of the Low A-S subjects to the aggression arousing situation deserves attention in its own right. Low A-S subjects in the Arousal group directed significantly less aggression toward the Jewish-named characters than did their matched partners in the Control group. Thus they showed a definite inhibition of postarousal fantasy aggression with respect to Jewish-named characters. Even more noteworthy is the fact that as a function of aggression arousal Low A-S subjects directed significantly less aggression toward Jewish-named characters than they did toward the non-Jewish characters. This differential response to the fantasy characters is the reciprocal of that found in the case of High A-S subjects. In effect, the tolerant subjects gave evidence of a specific tendency to avoid expressing aroused aggression toward Jews.

This finding challenges the traditional image of the very tolerant individual as one who is a bias-free neutral in his reactions to minority group members. The Low A-S subjects clearly demonstrated a response bias when angered that was consonant with their expressed attitudes of extreme tolerance. The finding agrees with the proposition that extreme attitudes in any direction are likely to entail an appreciable emotional involvement. True emotional neutrality, in the case of prejudice as with any other dimension of attitude, may be found most often in persons who lie between the positive and negative extremes.

[4] The procedure of comparing High and Low A-S subjects in terms of A-C differences insures that any significant differences between them are a function of aggression arousal. It is of interest, nevertheless, to examine the data for possible differences in fantasy aggression between High and Low A-S subjects under control conditions. An analysis of variance of the fantasy data obtained from the Control group indicated that the High and Low A-S subjects who were not subjected to the aggression arousal situation did not differ significantly in the amount of fantasy aggression they directed toward the four fantasy characters.

SUMMARY

In order to test a hypothesis derived from the scapegoat theory of prejudice, two groups of male college students, chosen on the basis of extremely high and extremely low scores on the Levinson Anti-Semitism Scale, were subjected to an aggression arousing situation. Following aggression arousal a specially designed fantasy test was administered, which yielded measures of aggression expressed toward fantasy characters with Jewish or non-Jewish names. Control subjects, matched with Arousal group subjects on the basis of A-S scale scores, took the fantasy test without prior aggression arousal.

Aggression arousal produced significant differences in the fantasy aggression of High and Low A-S subjects. The nature of these differences supported the hypothesis that persons high in anti-Semitism have a greater specific tendency than persons low in anti-Semitism to displace aggression selectively onto Jews. In addition, evidence was obtained suggesting that individuals very low in anti-Semitism have a specific tendency to avoid expressing aroused aggression toward Jews.

REFERENCES

BERKOWITZ, L. Anti-Semitism and the displacement of aggression. *J. abnorm. soc. Psychol.*, 1959, **59**, 182–187.

DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press., 1939.

LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296–309.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

WEATHERLEY, D. A. An investigation of the relationship between anti-Semitism and displacement, projection, and inhibition of aggression. Unpublished doctoral dissertation, Stanford University, 1956.

(Received February 1, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 458-460

SOME COMMENTS ON THE EPPS EGO DISJUNCTION SCORE

EDWARD F. GOCKA[1] AND VITALI ROZYNKO[2]
Veterans Administration Hospital, American Lake, Washington

An ego disjunction score (EDS) derived from four pairs of needs in the Edwards Personal Preference Schedule (EPPS) was reported to relate positively, in general, to degree of psychopathology (Trehub, 1959). The groups tested could be ranked on their mean EDSs, from low to high, in the following order: college students, adolescents, neurotics, character disorders, and schizophrenics. A statistical analysis yielded three significantly different subgroups. The first subgroup, with the lowest mean score, consisted of the college students alone. The second consisted of adolescents, neurotics, and character disorders, whereas the third subgroup, with the highest mean score, consisted of schizophrenics alone. The score did not discriminate at a significant level within the second subgroup. Several comments need to be made concerning this score.

The need pairs selected by Trehub (1959) were Aggression-Deference, Succorance-Nurturance, Autonomy-Abasement, and Order-Change. An ego disjunction indicator for a particular need pair was obtained by summing the standard score units for each need. The value of 100 was then subtracted from this sum, and positive residuals were considered as indicating the presence of disjunction. The degree of disjunction for each need pair was considered to be reflected in the magnitude of the positive residual. Thus, the total EDS for each subject consisted of the sum of his positive residuals on the disjunction indicators.

If we are to make comparisons between disparate classifications of people in terms of some personality score based on standard units, then we must be certain that contingency factors involving variables such as sex, age, education level, occupational status, and others of a demographic nature are controlled or accounted for in the analysis. Failure to control such contingency variables could lead one to impute some exotic discriminating power to a measure when, in reality, it was contingent on an ordinary demographic characteristic.

Trehub (1959) used the standard score distribution of the college normative sample published by Edwards (1954) in order to compute his EDSs. That these norms differ from the norms for an adult sample is evident from results obtained by Koponen (1957) and published by Edwards (1959) in the revised edition of the EPPS manual. If we now consider only the demographic differences within a particular group like the general adult population, we also expect to find these demographic differences reflected in the EPPS scores. Some data reported by Koponen (1957) for 3,882 male household heads illustrate this point. When Koponen divided his subjects into four age groups, significant differences (.001 level) between means were found for 13 of the 15 EPPS needs. Similarly, when this sample was divided in terms of grade school, high school, or college educational levels, 11 of the 15 EPPS needs gave significant differences at the .001 level. When the sample was divided into four income groups, 10 of the 15 EPPS needs gave significant differences at the .001 level. When divided by region of the country (east, central, south, west), 10 of the 15 needs gave the same general level of differences. Similar results occurred with other demographic variables and in a female sample numbering 4,922 subjects.

In this critique, we will focus only on the relationship of age to EDS in order to demonstrate our point. Before turning to the analysis of age effects, from which we may roughly predict our EDS increments for groups of different ages, we will consider some specific facts relating Koponen's general adult population to the college normative sample. If we consider differences in age, we find that 67.8% of Koponen's male group is of age 40 or above; in contrast the college normative sample has only 0.9% of its male subjects with an age of 40 or above. For the females, the percentages are 67.3 and 2.4, respectively. When considering Koponen's overall data, especially for those needs making up an EDS, we find that six of the eight mean need values for the general adult male population are higher than the mean need values for the college normative sample; the average raw score increase over the college normative group equals 2.44 per ego disjunction pair. Likewise, the average raw score increase over the college normative group per ego disjunction pair for the general adult female population equals 2.50. This is because five of the eight mean need values are higher for the general adult female population. The differences in standard deviations for the two populations appear to have neither the consistency nor the scope shown by their population means. The general adult male population standard deviations are

[1] The authors wish to acknowledge the assistance of Hildegund Holloway in data analysis and John B. Marks, chief research psychologist at American Lake, for his constructive review of the manuscript.

[2] Now at Stockton State Hospital, Stockton, California.

larger than those for the college subjects by an average of only .32 per ego disjunction pair. The general adult female standard deviations, on the other hand, are smaller than the college normative sample by an average of .04 per ego disjunction pair. At best, for the present samples, we can say that the differences between standard deviations are so small that neither group would be favored greatly toward a larger mean EDS from the nature of these differences.

From the foregoing considerations we may conclude that both the male and female adult populations would receive a higher average EDS than would the college normative sample simply because of the average increase in the values of the means for all ego disjunction pairs for the adult population. Indeed, if we take the raw means for each group and compute EDSs based on the parameters of the college normative sample, we obtain a measure of the expected increase in the average EDS for the particular group compared with the normative sample. That is, if we use the raw score means of any sample as our best estimate of a raw score in that sample, then our EDS computed from these means will give us some estimate of how much the actual average EDS will differ from the actual average EDS computed for the normative sample. Of course, if we use the means of the normative sample to compute its EDS we get a score of exactly zero. An actual average EDS for this group, based on the raw scores themselves, will always be greater than zero, however, due to the nonlinear method of scoring. Hence, our estimated increase for the normative group will be zero, while the increase for any other sample will be influenced by the difference of its means from the mean of normative sample. In applying this technique to the general adult population, we found that the expected increase was 26 for the adult males and 25 for the adult females!

Since Koponen's data were broken into four age groups, as well as by sex and education, the expected increases in average EDSs were obtained for each of these categories. To obtain an average expected increase, the means given by Koponen were first standardized on the college normative parameters, and these standard scores were combined into an EDS. Table 1 gives these expected increases. For example, we would expect a difference of 14 in the average EDSs computed for the male college norms and the general adult male sample between the ages 30 to 39, provided that the standard scores for the general adult group were based on the parameters of the male college normative group. Table 1 shows the consistency of the relationship to age of these expected increases. One may also note that the smallest expected increase occurs for college educated "household heads" under 30 years of age, the group which is the closest to the college normative sample.

TABLE 1

EXPECTED AVERAGE EGO DISJUNCTION INCREASES OVER THE COLLEGE NORMATIVE SAMPLE FOR GROUPS DIVIDED BY AGE, SEX, AND EDUCATION

(From Mean EPPS Values Given by Koponen, 1957)

Group	Age			
	Under 30	30–39	40–54	55 and Over
Combined Total				
Male	13	14	26	35
Female	15	22	23	28
College Educated				
Male	9	10	17	25
Female	6	14	17	26
High School Educated				
Male	11	15	29	35
Female	18	24	27	34

TABLE 2

COMPARISON OF EDSs FOR TWO AGE GROUPS

Age	Males		Females	
	<40	>40	<40	>40
Sample Size	13	43	34	66
Mean Age	34	55	34	54
Mean EDS	25.54	41.00	27.59	36.62
t value	2.83		2.47	
Probability Level	.01		.05	
Predicted Difference	17.00		7.00	
Actual Difference	15.46		9.03	

In order to verify the results of Table 1, a sample of voting age adults from a single precinct of the city of Seattle was divided by sex and age, and the actual EDSs were computed for each individual in the manner outlined by Trehub (1959). Since the number of subjects was not large, the groups were divided into those over 40 and those under 40. Differences between means were tested for significance for both the males and the females. A predicted difference between the age groupings was also computed from the data of Table 1. For example, to get the predicted difference between the two age groups for the males, the expected increases for the Combined Total-Male group were obtained as follows: $\frac{1}{2}(26 + 35 - 13 - 14) = 17$. Table 2 gives the pertinent information on this sample as well as a comparison with the expected results anticipated from Table 1. The fact that the EDS is related to age seems to be quite adequately demonstrated.

The demonstration that the age factor is important in EDSs still leaves some questions un-

answered. Trehub's (1959) report indicates that the differences between the schizophrenic and the neurotic and character disorder groups cannot be accounted for by age effects because these groups are essentially in the same age range. That other demographic effects, however, could have entered into the results is a speculation raised by Fordyce and Crow (1960), who failed to replicate a portion of Trehub's results on a group of adult subjects from the state of Washington. In contrast to Trehub's original study, the EDS failed to discriminate among neurotics, character disorders, acute schizophrenics, and chronic schizophrenics.

In conclusion, we can state that experimenters using the EDS must control such contigency variables as sex, education, and age. Other factors covered in Koponen's (1957) demographic study such as region of the country, rural-urban location, and income should also be considered. Care must also be exercised because of the nonlinear nature of some of these effects—for example, the higher EDS for the adolescents relative to the college students in Trehub's (1959) study. Only after all of these considerations are acted upon may a value of EDS become an unconfounded measure of intrapsychic conflict.

REFERENCES

EDWARDS, A. L. *Manual, Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1954.

EDWARDS, A. L. *Manual for Edward's Personal Preference Schedule.* (Rev. ed.) New York: Psychological Corporation, 1959.

FORDYCE, W. E., & CROW, W. R. Ego disjunction: A failure to replicate Trehub's results. *J. abnorm. soc. Psychol.*, 1960, **60**, 446–448.

KOPONEN, A. The influence of demographic factors on responses to the Edwards Personal Preference Schedule. Unpublished doctoral dissertation, Teachers College, Columbia University, 1957.

TREHUB, A. Ego disjunction and psychopathology. *J. abnorm. soc. Psychol.*, 1959, **58**, 191–194.

(Received February 26, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 461-464

EFFECTS OF AGGRESSIVE CARTOONS ON CHILDREN'S AGGRESSIVE PLAY

PAUL MUSSEN AND ELDRED RUTHERFORD[1]
University of California, Berkeley

DESPITE widespread concern about the possible harmful consequences of exposure to the violence and aggression prominently portrayed in movies, television, and comic books, there has been relatively little systematic research on the problem. Two diametrically opposed hypotheses about the effects have been proposed: one emphasizing the possible beneficial results, the other stressing the possible deleterious effects. According to the often-invoked catharsis hypothesis, involvement in fantasy aggression may serve as a "displacement," providing a harmless "release" for children's hostile impulses, and thus reducing the instigation to overt acts of aggression (Dollard, Doob, Miller, Mowrer, & Sears, 1939). On the other hand, a number of popular reports, based on the experiences of criminologists and psychiatrists who have worked with criminals and delinquents (e.g., Wertham, 1953), maintain that the portrayal of violence in the mass media stimulates hostile impulses and increases antisocial behavior.

Experimental confirmation of the catharsis hypothesis applied to fantasy aggression was obtained by Feshbach who found that "*self-initiated* verbal responses" (italics ours), writing aggressive TAT stories following the arousal of aggression, led to a reduction in the amount of subsequent interpersonal aggression manifested (Feshbach, 1955). The results of other studies of children's reactions in play situations, however, are clearly inconsistent with this finding (Feshbach, 1956; Siegel, 1956). For example, in a later study, Feshbach (1956) found that second and third grade children who played with aggressive toys displayed significantly *more* inappropriate aggression (aggression not directly determined by the type of play object, e.g., striking or insulting another child) than those who played with neutral playthings. Similarly, Siegel (1956) found that preschool subjects tended to be more aggressive in social interactions with peers after viewing an aggressive animated cartoon than they were after viewing a neutral one, although the difference was not statistically significant.

On the basis of these findings, it may be hypothesized that under permissive conditions, indulgence in aggressive fantasy may result in increased aggressive play. Such an increase may be the consequence of identification with aggressive fantasy characters, or, under these conditions, fantasy-produced reduction of inhibition against the expression of aggression, or both.

The present research was designed to test the specific prediction, derived from this hypothesis, that exposure to violence in an animated cartoon may intensify the aggression that children manifest in their play activities. The possible effects of two other variables were also investigated: experimentally-induced frustration, since both Feshbach and Siegel suggest that this may strongly influence reactions to aggressive fantasy; and sex of subject, since boys and girls may react differently to this kind of stimulus.

PROCEDURE

A 3 × 2 × 2 factorial design was employed in this investigation. The subjects, 36 lower middle class first grade pupils between the ages of 6 and 7, were randomly assigned to one of six groups, each group consisting of three boys and three girls. Three groups, the *frustration groups*, were assigned a task which, according to their teacher, was both tedious and frustrating to the children. It was similar to the frustration experience used by Yarrow (1948) and consisted of copying numbers repeatedly while the teacher criticized their work frequently. Immediately following the frustration session, one group viewed an aggressive cartoon, another viewed a nonaggressive cartoon, and the third, no cartoon. Similarly, each of the other three groups was assigned to one of the three treatment conditions—viewing an aggressive cartoon, a nonaggressive cartoon, or no cartoon—but none of these groups were experimentally frustrated beforehand.

The cartoons. The aggressive cartoon showed a continuous sequence of aggressive activity in which an animated weed attempted to choke a flower, and a panda bear struggled to destroy the weed. The theme of the nonaggressive cartoon was the fun of playing cooperatively, a frog and duck being the central characters portrayed. Each cartoon was exactly 8-minutes long.

Immediately after seeing the cartoons, the children were taken individually to a room where the test of aggressive drive was administered by one of the six young women who acted as testers. The two groups of children who did not see a cartoon were taken directly from their classroom to the testing room.

The test of aggression. The strength of aggressive drive must, of course, be inferred from behavior. Siegel's indices were based on overt aggression in an interpersonal situation with peers (Siegel, 1956). The measure used in the present study was the verbal expression of desire to destroy an inanimate object, a balloon. The test used, adapted from Stone's (1956) balloon-breaking procedures, consisted of eight questions,

[1] The authors gratefully acknowledge the assistance of Maryly Blew, a first grade teacher in Berkeley elementary school system, in conducting this study.

asked orally in fixed sequence, about an inflated balloon held by the tester. Five of the questions concerned the child's desire to "pop" the balloon himself to or see the tester do it (e.g., "Do you think it would be fun to see me 'pop' this balloon?"; "If I gave one to you, would you take it home and 'pop' it, or would you just play with it?"). Subjects were not actually given an opportunity to break the balloon, however. The child's aggression score was simply the number of times he chose the "popping" response. Although the possible range of scores was 0–5, the actual range was 0–2.

The testers, who did not know anything about the experimental treatments the children had experienced prior to the test, were also asked to check any of the following adjectives that they felt described the subject's behavior during the testing period: angry, excited, tense, relaxed, cooperative. At the conclusion of the experiment, all children who had been frustrated were given balloons to take home and permitted to see another, nonaggressive cartoon as rewards for their cooperation.

RESULTS

The results of analysis of variance of the aggression test data, summarized in Table 1, indicate that the only significant effect was the main effect for the cartoon variable. Table 2, which gives the mean aggression scores of the six groups, shows that subjects who saw the aggressive cartoon were more willing to express aggressive impulses in this setting than those who saw the nonaggressive cartoon or no cartoon at all. This finding supports the hypothesis that viewing violence in a cartoon may actually stimulate or intensify the child's aggression in a subsequent permissive situation. There was no evidence in these data that the sexes differed from each other in their reactions to the cartoon.

TABLE 1
ANALYSIS OF VARIANCE FOR AGGRESSION-TEST DATA

Source	*df*	*SS*	*MS*	*F*	*P*
Cartoon	2	11.05	5.51	15.31	< .01
Frustration	1	.44	.44	1.22	NS
Sex	1	1.00	1.00	2.78	NS
Frustration × Sex	1	0	0		
Frustration × Cartoon	2	.08	.04		
Sex × Cartoon	2	.52	.26		
Frustration × Sex × Cartoon	2	.11	.06		
Within cells	24	8.69	.36		

TABLE 2
GROUP MEANS FOR AGGRESSION-TEST DATA

Cartoon Treatment	Frustrated	Nonfrustrated
Aggressive Cartoon	1.67	2.00
Nonaggressive Cartoon	.67	.83
No Cartoon	.50	.67

It is interesting to note that, contrary to what would be predicted on the basis of the frustration-aggression hypothesis, the frustrated and nonfrustrated groups experiencing the same cartoon treatments did not differ significantly with respect to aggressive expression as measured here. The data from the adjective check lists, filled in by the testers, indicated, however, that feelings of anxiety and tension were more likely to occur subsequent to seeing an aggressive cartoon than after the other treatments. Eight of the 12 subjects (67%) who saw the aggressive cartoon were described as "tense" while among the 24 other subjects, only 8 (33%), 4 who had seen the nonaggressive cartoon and 4 who had not seen any cartoon, were rated in this way ($\chi^2 = 3.94$, $p < .05$).

These findings may be interpreted as evidence that reaction to frustration did not increase overt aggression but rather heightened feelings of tension—i.e., that the frustration-aggression hypothesis did not hold in this situation. On the other hand, it is also possible that, in spite of the rationale underlying the use of this frustration procedure, the task assigned to the "frustration" group was not actually frustrating. In view of these considerations, it is difficult to make any generalizations about the possible effects of experimentally-induced frustration or subsequent aggression.

DISCUSSION

The experimental results provide evidence supporting the basic hypothesis of the study which stated that film-mediated fantasy aggression may intensify aggressive expression in permissive play. Since the present findings may be directly applicable only to the restricted situation of permissible aggressive play, it is difficult to evaluate their implications, if any, for the validity of the catharsis hypothesis. This hypothesis is generally interpreted, as it was in the Feshbach and Siegel studies, to refer to *interpersonal* aggression subsequent to other aggressive activities, such as a fantasy aggression. There is no evidence in the data to suggest that, in an unselected group of subjects, those who increase in the measure of aggression used here become more hostile toward other children, although pretests show that this measure differentiated at the .02 level the children rated high and low in aggression by their nursery school teachers. Nevertheless, it should be noted that our findings are entirely consistent with the results of Feshbach's second experiment in which children's "inappropriate aggression" increased during play with aggressive toys. If, as is generally assumed, a child's aggression is largely a function of the strength of his aggressive drive

and his internalized inhibitions against aggression, then increases in aggression must be due to either temporary intensification of drive or reduction of inhibitions, or both. In the present study, intensification of instigation to aggression may have resulted directly from identification with aggressive cartoon characters with accompanying assumption of their motives. In Feshbach's study, the aggressive toys may have provided cues eliciting associations with real or fictional aggressive identificands, and consequently, increased aggressive motivation.

The second possible, perhaps more parsimonious, explanation is that the permissive "fun" context of animated cartoons or play with aggressive toys leads to a reduction of inhibitions against aggressive expression, particularly in situations where aggressive play is not likely to be punished. It has frequently been observed that aggression in doll play increases from session to session, presumably because "the deliberate permissiveness of the experimental procedure progressively acts to reduce the inhibitions about aggression the child has heretofore acquired" (Levin & Sears, 1956, p. 149). Moreover, Feshbach (1956) found that, particularly among boys initially low in aggression, aggressive behavior increased following permissive free play experiences. The child may regard these situations as "permission-giving," with respect to aggression (Siegel & Kohn, 1959); i.e., he may infer that aggression is distinctly permissible and appropriate in these situations. The response of behaving aggressively may then generalize to other relatively permissive situations, e.g., doll play or a balloon-breaking test.

These considerations may also help to explain the discrepancy between the findings of the present study and those of Siegel who found that children did not display significantly more aggression toward peers after watching an aggressive cartoon than they did after watching a control cartoon. The two studies appear to differ most importantly in the indices of aggression employed.[2] In the test used in the present study, a relaxed tester asked questions phrased in a way that permitted, or even encouraged, the free expression of aggressive feeling, not overt aggression. The test setting contained no apparent threat of counteraggression and there were no strong cues evoking internalized inhibitions against aggression. Increased aggressive feelings, stimulated by the aggressive cartoon, were therefore likely to be generalized and expressed in a permissive test situation.

Siegel's measure of aggression, on the other hand, was based on overt aggression in social play in the absence of adults. This is a situation in which strong inhibitions against aggression are likely to be evoked (Siegel & Kohn, 1959). Siegel and Kohn (1959) have pointed out that "lacking any adults to define the social situation and to express expectations of his behavior in it, the child will . . . rely . . . on his own learned standards of conduct . . . which, in the middle-class child, will typically be that aggression is unacceptable in social play" (p. 134). It may be reasonably hypothesized that there was a reduction of inhibition against aggression among Siegel's subjects, as among ours, after they viewed an aggressive cartoon, but this effect was not apparent in a test situation that contained cues that elicit powerful inhibitory responses.

[2] While Siegel's subjects were somewhat younger than ours, this age difference could hardly explain the discrepancy in the results. It is also possible that differences in the cartoon stimuli account, at least partially, for the differences in the findings. For example, compared with those used by Siegel, the aggressive cartoons of the present study may have provided better (i.e., more immediate or intimate) identification models or a more "permission-giving" context for aggression. There is no evidence to support such an explanation, however.

SUMMARY

The purpose of this study was to test the hypothesis that exposure to aggressive fantasy in an animated cartoon may intensify children's impulses to aggression. Subjects were 36 first grade children, 18 girls and 18 boys, of middle class origin. The intensity of the child's aggressive impulses was inferred from his responses to questions concerning desire to "play with" or "pop" a large yellow balloon held by a tester.

The experimental findings clearly supported the major prediction that exposure to aggressive fantasy in an animated cartoon would stimulate children's aggressive behavior in play. Two possible explanations for these results were discussed: that intensification of instigation to aggression may have resulted from the child's identification with the aggressive cartoon characters, with accompanying assumption of their motives; that the relaxed, "fun" context of animated cartoons may lead to a reduction of inhibitions against aggressive expression in a permissive play situation.

REFERENCES

DOLLARD, J., DOOB, W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression.* New Haven: Yale Univer. Press, 1939.

FESHBACH, S. The drive-reducing function of fantasy behavior. *J. abnorm. soc. Psychol.*, 1955, **50**, 3–11.

FESHBACH, S. The catharsis hypothesis and some consequences of interaction with aggressive and neutral play objects. *J. Pers.*, 1956, **24**, 449–462.

LEVIN, H., & SEARS, R. R. Identification with parents as a determinant of doll play aggression. *Child Develpm.*, 1956, **27**, 135–153.

SIEGEL, ALBERTA E. Film-mediated fantasy aggression and strength of aggressive drive. *Child Develpm.*, 1956, **27**, 365–378.

SIEGEL, ALBERTA E., & KOHN, L. G. Permissiveness, permission, and aggression: The effect of adult presence or absence on aggression in children's play. *Child Develpm.*, 1959, **30**, 131–141.

STONE, L. J. Aggression and destruction games: Balloons. In L. B. Murphy (Ed.), *Personality in young children*. New York: Basic Books, 1956. Ch. 10.

WERTHAM, F. *Seduction of the innocent*. New York: Rinehart, 1953.

YARROW, L. J. The effect of antecedent frustration on projective play. *Psychol. Monogr.*, 1948, **62**(6, Whole No. 293).

(Received March 2, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 465–467

RELATIONSHIPS BETWEEN HOPE OF SUCCESS AND FEAR OF FAILURE, ANXIETY, AND NEED FOR ACHIEVEMENT

E. EDWARD REITMAN[1] AND CARL D. WILLIAMS[2]

University of Miami

An issue regarding the nature of what is measured by McClelland's TAT test of need for achievement (n Achievement) has developed from a study by Clark, Teevan, and Ricciuti (1956). On the basis of a level-of-aspiration questionnaire, they placed the subjects, on whom they had obtained an Achievement score, along a continuum from hope of success (HS) to fear of failure (FF). They report that, as compared to the middle third of the continuum, the HS and FF groups had significantly lower n Achievement scores and significantly more positive subscores and fewer negative subscores on the n Achievement measure. The first finding of Clark et al. (1956) is not in agreement with previous work summarized in McClelland, Atkinson, Clark, and Lowell (1953). They stated that the n Achievement score could not be used as an indication of the HS-FF continuum and attributed their discrepancy in results to the difference in the pictures used to elicit the stories.

By relating their findings to McClelland's (McClelland et al., 1953) formulation, Clark et al. (1956) indicated that their HS group would be expected to have had little anxiety, the middle group some anxiety, and the FF group great anxiety over goal attainment. After thus suggesting that the HS and FF groups were *genotypically* different in regard to anxiety over goal attainment, they indicated that the two groups were *phenotypically* similar as measured by McClelland's n Achievement test.

A directly opposite interpretation of these results has been suggested by Atkinson (1958). He regards the TAT score of n Achievement as a measurement of genotypic and the level-of-aspiration questionnaire as an expression of phenotypic motivation.

The measurement of anxiety, in addition to level-of-aspiration and n Achievement, may permit the issue to be resolved. It would follow that if the proposal by Clark et al. (1956) is correct, the FF group should be more anxious than the middle group, which in turn should be more anxious than the HS group. If Atkinson (1958) is correct in regarding both the HS and the FF groups as being genotypically fearful of failure, then these groups should be more anxious than the middle group but should not differ in anxiety between themselves.

The present investigation had three objectives: first, to attempt to clarify the contradictory results by repeating the Clark et al. (1956) study using pictures similar to those used in the earlier studies; second, to explore the generality of the Clark et al. findings by employing two additional samples of subjects; third, to contribute to the understanding of what is measured by McClelland's TAT measure of need for achievement and level-of-aspiration.

METHOD

The 113 subjects in this research were male students attending the University of Miami summer sessions. Seventy-two were enrolled in two introductory psychology courses taught during one semester by the same professor, using identical textbooks and examinations. Forty-one students were in one of three more advanced undergraduate psychology courses taught by three different professors.

The procedure followed was as reported by Clark et al. (1956). The group TAT measure of n Achievement was administered early in the semester. In all, five individual testing sessions were required in order to administer the TAT to the 113 subjects. This was done under neutral conditions using the same instructions, experimenter, and physical location with the following exceptions. One of the introductory groups was not tested in the same room and a new experimenter administered the TAT during a testing session for one of the advanced groups. Statistical tests of the TAT scores showed no significant difference between means of the two introductory groups or between means of the three advanced groups. The four pictures used to elicit the stories were obtained from McClelland,[3] who stated that they were "roughly equivalent" to those used in the original studies. The pictures are numbered 33, 26, 24, and 9 in Atkinson (1958).

The TAT protocols were independently scored by two persons having validity coefficients of .85 and .88 with regard to the material included in the manual provided by Atkinson (1958). On the present set of scores, a rho of .87 was obtained between the ratings of these same persons. Later in the semester, the subjects were given the Taylor Manifest Anxiety (*MA*) scale.

The level-of-aspiration questionnaire developed by

[1] Now at the University of Houston.

[2] Both authors shared equally in designing the study and in preparing the report. Edward Reitman collected the data, scored the tests, and made the statistical analyses. The authors thank Charles F. Williams for also scoring the TAT protocols.

[3] Personal communication.

Clark et al. (1956) was administered immediately before the final examination. Again the analysis of the data obtained from the level-of-aspiration scale was similar to that used by Clark et al. In each case the range of the raw scores for each index was converted into seven class intervals and the value of seven was assigned in such a manner so as to insure that the HS subjects always received the highest scores.

RESULTS

Table 1 presents the intercorrelations between the four indices of the questionnaire for the introductory students and for the more advanced students. The correlations for the advanced group are very similar to those reported by Clark et al. (1956). The correlations for the introductory group are somewhat higher than those for the advanced group, very probably indicating the greater degree of homogeneity in this group. Both sets of correlations are positive and highly significant indicating the internal consistency of these indices. The four indices were summed and the resulting distribution divided into three parts. The upper third constituted the hope of success group and the lower third the fear of failure group.

Table 2 presents the mean n Achievement score for the various groups. The mean for the introductory students is significantly higher ($p < .01$) than that for the advanced students. None of the six F tests between the HS, middle, and FF categories were significant.

Table 3 shows the mean MA scale scores for the HS, middle, and FF categories. For both the introductory and advanced students the mean MA scale score increases from the HS group to the FF group. In neither case does this trend reach significance. However, for the advanced students, the difference between the FF group and the HS and middle groups combined is significant ($p < .05$).

Two additional analyses reported by Clark et al. (1956) were also made on the present data. However, the relationships between the positive and negative n Achievement subscores for the HS, middle, and FF groups, and the median analysis of HS-FF in relation to frequency of positive and negative categories which were presented in Clark et al.'s Tables 4 and 5 have been eliminated since none of the comparisons proved statistically significant.

TABLE 1

RELATIONSHIP AMONG FOUR INDICES OF HOPE OF SUCCESS AND FEAR OF FAILURE

(Each index is correlated with the sum of the other three)

Index	Correlation with Three Remaining Indices[a]	
	Introductory Group	Advanced Group
Upper Limit minus Settle For	.90	.82
Upper Limit minus Settle for Range	.93	.86
Affect over Settle For	.94	.85
Affect Upper Limit minus Affect Settle For	.91	.80

[a] All of the correlations are significant beyond the .01 level.

TABLE 2

TOTAL n ACHIEVEMENT SCORES FOR THE HOPE OF SUCCESS, MIDDLE, AND FEAR OF FAILURE GROUPS

(F_1 = Between groups/within groups; F_2 = Middle and combined HS, FF group/within groups; F_3 = FF and combined HS, Middle group/within groups)

Group	Introductory Group		Advanced Group	
	N	Mean n Achievement Score	N	Mean n Achievement Score
HS	20	6.15	15	2.47
Middle	28	5.21	12	2.50
FF	24	7.35	14	.79
Totals	72	6.18 F_1, F_2, F_3-ns	41	1.91 F_1, F_2, F_3-ns

TABLE 3

TOTAL MA SCALE SCORES FOR THE HOPE OF SUCCESS, MIDDLE, AND FEAR OF FAILURE GROUPS

(F_1 = Between groups/within groups; F_2 = Middle and combined HS, FF group/within groups; F_3 = FF and combined HS, Middle group/within groups)

Group	Introductory Group		Advanced Group	
	N	Mean MA Scale Score	N	Mean MA Scale Score
HS	20	12.85	15	10.87
Middle	28	13.11	12	11.16
FF	24	15.42	14	15.36
	F_1, ns F_2, ns	F_3, ns	F_1, ns F_2, ns	$F_3 = 4.54$ $p = .05$

DISCUSSION

The intercorrelations obtained among the indices of the level-of-aspiration questionnaire, although slightly higher, closely parallel those of Clark et al. (1956). These results are not too surprising since individuals within both groups were quite homogeneous and, moreover, the very nature of the correlated items themselves tends to suggest some expected degree of relationship.

It might be noted from Table 2 that both sets of protocols, although similar in length and scored by the same judges, differed markedly in their

mean n Achievement scores. However, the TAT scores for the introductory psychology students were significantly higher than for the more advanced psychology students. Any of several factors may have been responsible for the difference. First, the sets of protocols, although scored independently by the same two judges, were scored at different times. However, the high degree of correlation between their ratings makes this alternative seem doubtful. More probably, the nature of the two groups may have contributed to the result. We may assume the more advanced students to be more sophisticated and therefore more guarded in their answers.

Contrary to Clark et al.'s (1956) data and to the early work summarized in McClelland et al. (1953), the present results do not reveal any relationship between the HS-FF continuum and n Achievement scores. The fact that the pictures used in the present study were roughly equivalent to those in the earliest studies strongly suggests that the different results among the various studies are probably due to variables other than or in addition to the pictures used. Thus the present results lend strong support to the statement by Clark et al. (1956) that "the level of n Achievement score *cannot* be used as an index of the FF-HS continuum" (p. 186).

The *MA* scale was administered in an attempt to resolve the issue between Clark et al. (1956) and Atkinson (1958) concerning the genotypic vs. phenotypic status of the level-of-aspiration questionnaire and McClelland's TAT test, as measures of motivation. Analysis of the relationship between *MA* scale scores and the HS-FF continuum tend to support the position of Clark et al. that the HS and FF groups, as measured by a level-of-aspiration questionnaire, differ in level of anxiety. It was found that the higher the hope of success, the less the anxiety; and conversely, the greater the fear of failure, the greater the anxiety. Therefore Atkinson's view that the FF and HS groups are only phenotypically differentiated and both are characterized by underlying basic fear of failure is not supported by these results.

SUMMARY

Previous research has raised a question as to whether McClelland's TAT test of n Achievement measures genotypic or phenotypic motivation. The present investigation measured level-of-aspiration, manifest anxiety, and need for achievement in college students in an attempt to resolve the issue. The major findings follow:

1. Students in introductory psychology classes made significantly higher n Achievement scores than students in more advanced psychology classes.
2. There were no differences in n Achievement among groups characterized as being hopeful of success, middle, or fearful of failure.
3. For students in advanced psychology classes, those fearful of failure showed significantly more manifest anxiety than those in the middle or hope of success groups combined.

Generally these results showed no relationship between the HS-FF continuum and TAT scores, introduced level of sophistication of the subjects as a variable that seems to affect n Achievement level, and did not support the view that McClelland's TAT test of need for achievement measures basic or genotypic motivation.

REFERENCES

ATKINSON, J. W. (Ed.) *Motives in fantasy, action, and society*. New York: Van Nostrand, 1958.

CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope of success and fear of failure as aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, **53**, 182–186.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

(Received March 10, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 468–471

THE RELATIONSHIP BETWEEN ANXIETY AND *ORAL* WORD ASSOCIATION PERFORMANCE[1]

MICHAEL J. GOLDSTEIN
University of California, Los Angeles

SARASON (1959) has recently presented evidence that subjects differing in anxiety level respond differentially to a group Word Association test. His highly anxious subjects tended to respond with greater variability to the same stimulus items on their second administration and tended to give a greater number of rare responses, according to the Minnesota norms (Russell & Jenkins, 1954), than did subjects lower in anxiety. These trends were markedly increased under conditions of ego threat in which the subject was informed that the Word Association test was a sensitive measure of personality. Sarason concluded that these results support the habit or interfering-response theory of anxiety rather than a drive theory (Spence, 1958; Taylor, 1958).

Before one can accept such a conclusion, however, it is necessary to determine whether these findings are not also congruent with a drive theory. In the theory explicated recently by Taylor (1958), the relationship between anxiety and performance is conditional upon the structure of the subject's habit hierarchy upon entering the experimental situation. If the response to be learned (or emitted, as in the Word Association test) is strong in the hierarchy and involves few competing responses, high anxious (HA) subjects should perform more rapidly or more efficiently than low anxious (LA) subjects. On the other hand, if there is no strong dominant response and many competing response tendencies of approximately equal strength, HA subjects should perform more slowly and less efficiently than their LA counterparts.

Both of these predictions are based on the Hullian notion that drive acts indiscriminately to raise the reaction potential of all habits in the hierarchy. Following this line of reasoning, it is difficult, without some prior knowledge of the subject's hierarchy of responses to the stimulus words on a Word Association test, to make any prediction from drive theory discriminable from one deduced from a habit or interference theory. If the words used by Sarason (1959) elicited a number of competing verbal responses with none dominant over the others, then his finding of greater discrepancy in responses from one administration of the test to the others is in keeping with a drive theory of anxiety. Even the enhancing effects of experimental instructions could be accounted for by drive theory if such instructions are viewed as raising the subject's drive level. If, on the other hand, the words used by Sarason consisted of stimulus items with strongly associated responses (i.e., Black-White) and only minimally competing responses, then Sarason's results support a habit theory of anxiety.

The present study is concerned with examining the relationship between anxiety and word association responses with the subject's habit hierarchy taken into consideration. Laffal (1955) has suggested that the number of different responses (*D*) given to a stimulus word by an aggregate of subjects may provide some crude statistical estimate of the habit hierarchy for each individual subject. For example, the word Boy tends to elicit few different responses in a group of subjects; most respond with Girl. According to Laffal, we can infer from such group data that the stimulus Boy is associated with a single, dominant response for most subjects and that there are few competing responses present. On the other hand, the word Flatter tends to elicit almost as many different responses as there are subjects responding to it. This kind of stimulus word has no dominant response associated with it, but evokes a number of competing responses of low strength in each subject. Laffal marshals some impressive data on reaction times and number of reproduction faults to support his inference that group observations of the number of different responses reflect individual habit hierarchies. According to drive theory, the relationship between anxiety and word association performance is a direct function of the *D* value of the stimulus word (number of different responses stimulated by the word). For low *D* words, high levels of anxiety should result in shorter reaction times, fewer reproduction disturbances, and fewer rare responses than lower levels of anxiety. For high *D* words, the reverse prediction applies. If the drive theory of anxiety is correct, we should find a significant interaction between anxiety and *D* level of the words, whereas a habit or interference theory predicts a significant main effect due to anxiety (regardless of the *D* level of the words) and no significant interaction. The present study attempted to examine these differential predictions.

[1] This study was carried out on a grant from the Faculty Research Committee at UCLA. The author would like to express his appreciation to Samuel Himmelfarb for running subjects and analyzing the data.

A second and less important aspect of this study involves the oral administration in individual sessions of word association material. Although group forms of such tests are valuable for studying verbal habits, the investigation of personality factors in verbal behavior requires proper attention to the possibility that untimed group forms may primarily reflect the subject's pace of working and his tendency towards censoring his responses as much as it does his verbal habits. For a more critical test of the drive vs. habit theories, the individual, orally-administered method was used in the present study.

METHOD

Subjects. Subjects were students in the introductory psychology course at UCLA who volunteered as part of their course requirement. Each subject was administered the 50-item form of the Heineman (1953) forced-choice version of the Taylor scale of anxiety (*HMA* scale) in class. From the total distribution of scores, 20 HA and 20 LA subjects were selected. Since a preliminary analysis of the scores revealed a significant sex difference, subjects were chosen from the distribution of their own sex. Each anxiety group contained 10 males and 10 females. HA males scored 20 or above on the *HMA* scale and HA females 24 or above; LA males scored 10 or below on the *HMA* scale, while LA females scored 7 or below.

Procedure. Approximately one month after the administration of the *HMA* scale, the classes were told that a random sample had been drawn, and the sampled people were asked to participate in an experiment. So far as could be determined, no subject was aware of a connection between his *HMA* scale score and his selection for the experiment. These subjects were given a 99-item Word Association test similar to that used by Laffal (1952), the only change being that 12 words were substituted for homophones in the Laffal list. The subject was instructed to respond to each stimulus word with the first word which came to mind. On the second administration, he was asked to give the same word which he gave on the first administration. From the total of 99 words, 10 were selected as high, and 10 as low *D* words on the basis of the number of different responses given by the total number of 40 subjects. Because this step involved the creation of norms on the very sample to which they were to be applied, a check was made at the stability of the *D* values by correlating those obtained here with the values for all overlapping words in the Laffal study. A correlation of .75 was found for the 87 words, suggesting considerable stability from sample to sample.

RESULTS

Reaction time data. Each subject's reaction time to the first administration of the Word Association test was recorded to the nearest tenth of a second. In order to correct these scores for skewness, a reciprocal transformation was applied to them (Edwards, 1950). Table 1 presents the means and standard deviations of the *raw* reaction times for HA and LA subjects as a function of the *D* level of the stimulus word. The trend for both *D* levels is for HA subjects to react more slowly than the LA subjects. Analysis of variance of the reciprocal reaction times revealed the anxiety variance to be nonsignificant ($F = 3.42$, $.10 < p > .05$), the variance for *D* level to be highly significant ($F = 25.25$, $< .001$), and the Anxiety × *D* interaction to be lacking in statistical significance.

TABLE 1

MEANS AND STANDARD DEVIATIONS OF RECIPROCAL OF REACTION TIMES BY ANXIETY AND *D* LEVEL OF THE STIMULUS WORD

	High *D* Words		Low *D* Words	
	M	*SD*	*M*	*SD*
High Anxiety				
Males	.455	.069	.907	.194
Females	.550	.072	.906	.133
Low Anxiety				
Males	.560	.131	1.020	.211
Females	.580	.162	.957	.113

Reproduction faults. A count was made of the proportion of times the subject failed to give the same response to the same stimulus word on the second administration. Table 2 presents the means and standard deviations of these reproduction discrepancy scores by anxiety and *D* level. Analysis of variance yielded significance only for *D* level ($F = 95.01$, $p < .001$). The Anxiety × *D* interaction was not significant, and an inspection of Table 2 reveals that LA subjects tended to react with a larger number of reproduction faults to the high *D* words, contrary to both a drive or a habit interference theory.

Communality of response words. Sarason (1959) notes that anxiety level relates positively to the number of rare responses emitted. No norms comparable to those used by Sarason were available, but an analogous method of measuring communality of response was used, taking the *D* level of a word as an index. If a word tends to be responded to with a small number of dominant responses by the total group (a low *D* word), then we can determine whether HA subjects give more different responses from LA subjects. The same analysis can be done for high *D* words. The means and standard deviations of the proportion of different responses given by each anxiety group are presented in Table 3. With 10 subjects in each subgroup, it was possible for the group to give from 0–10 different responses to each stimulus word. Analysis of variance failed to reveal any significant differences except for *D* level ($F = 365.55$, $p < .001$, for 1 and 72 *df*).

TABLE 2

MEANS AND STANDARD DEVIATIONS OF THE PERCENTAGE OF REPRODUCTION FAULTS BY ANXIETY AND *D* LEVEL OF THE STIMULUS WORDS

	High *D* Words		Low *D* Words	
	M	*SD*	*M*	*SD*
High Anxiety				
Males	3.2	.98	1.3	.90
Females	3.3	1.19	.8	.75
Low Anxiety				
Males	3.6	1.74	.6	.66
Females	3.5	2.06	.8	.68

TABLE 3

MEANS AND STANDARD DEVIATIONS OF THE PROPORTION OF DIFFERENT RESPONSES GIVEN TO EACH STIMULUS WORD BY ANXIETY AND *D* LEVEL OF THE STIMULUS WORD

	High *D* Words		Low *D* Words	
	M	*SD*	*M*	*SD*
High Anxiety				
Males	.88	.077	.28	.141
Females	.91	.070	.37	.149
Low Anxiety				
Males	.90	.118	.29	.145
Females	.87	.118	.32	.166

Analysis over all 99 words. Although not as important theoretically, the same analysis was done over all 99 words on the word association task to test whether Sarason's (1959) results are tied to the group administered, untimed version of the Word Association test. It is possible that Sarason's results on anxiety reflect factors such as censorship of responses or differences in the pace of responding more than interference with verbal habits. An analysis of reciprocal reaction times, done on average RT over all 99 words, revealed a similar trend for *slower* reaction times for HA subjects (F for anxiety level = 2.914, $.10 < p > .05$ for 1 and 95 *df*) to that reported above. An analysis of variance on the total number of reproduction faults given by the subject to all 99 words, a measure which Sarason found related to anxiety level, showed no significance for anxiety level. A final analysis was done on the equivalent of Sarason's communality score described in the section above. This score, the number of different responses given to each stimulus word, also failed to reveal any relationship to anxiety ($F = 1.1135$, $p > .10$). These analyses fail to support Sarason's findings in any respect.

DISCUSSION

The present study, although failing to replicate Sarason's (1959) findings on the group Word Association test, do tend to support his conclusion concerning a habit interference theory of anxiety. The presence of slower reaction times to words with either simple or complex habit hierarchies is not compatible with drive theory and could indirectly be derived from a theory which assumes greater interfering response tendencies in HA individuals. This conclusion rests, however, on an assumption which warrants closer scrutiny.

The assumption is that the group data which yield *D* scores for stimulus words accurately reflect *individual* habit hierarchies. Some support for this assumption is provided from the reaction time and reproduction fault data, which indicate that we are dealing with two distinct classes of words. High *D* words *always* have longer reaction times than low *D* words and are much more likely to elicit response faults than low *D* words. Both findings are compatible with the idea that high *D* words do not have any strong dominant response associated with them and that the responses which are associated with them are strongly competing ones. The rapid reaction times and infrequent response faults to low *D* words suggests the reverse conclusion about the habit hierarchy associated with these words. These conclusions are supported more directly by a study by Rosen and Russell (1957) in which it was found that chains of associations given by individuals to repeated presentations of the same stimulus word parallel the cultural frequencies collated from group data. The assumption of different habit hierarchies associated with high *D* and low *D* does seem to be a reasonable one.

The failure to replicate Sarason's (1959) findings, then, is difficult to interpret. The present study involved a different form of the Taylor scale and a different group of stimulus words, factors which could account for the discrepancy. In spite of these differences in technique, however, it does seem possible that the factor of censorship over responses could most reasonably explain Sarason's findings. If HA subjects are more threatened by any testing situation than LA subjects, then they may be more likely to inhibit their responses and to censor them more carefully. These attempts to censor or inhibit responses could interfere with the recall of the first response on the second administration of the test. From this point of view, the present results are not incompatible with Sarason's. It is probably more difficult to censor or inhibit responses on the oral type of Word Association test. Nevertheless, the trend toward longer reaction times for HA over LA subjects may reflect an attempt in that direction. With the group administration, particularly under the untimed condition used by Sarason, such attempts at censorship could be more actively pursued, resulting in

greater discrepancy scores and a greater number of rare responses for HA subjects.

The results reported here suggest that under certain conditions, HA subjects emit more interfering response tendencies than LA subjects. But there are several studies, in which the drive theory of anxiety is clearly substantiated (Spence, Farber, & McFann, 1956; Taylor, 1958), which seriously question the validity of the dichotomous view in which the Taylor scale either measures drive level or interfering response tendencies. It probably measures both reaction tendencies, and the difficulty posed here and elsewhere is that there is no way of anticipating on a theoretical level when the interfering responses are prominent and when the enhanced drive is more likely to influence performance. Until a more explicit statement of the relation between drive level (D) and responses learned to the drive stimuli (s_D) is forthcoming, the predicted relationships between anxiety and performance will be frequently in error.

SUMMARY

The present study was concerned with the *oral* word association responses of subjects varying in anxiety level. Two classes of words, believed to vary in the simplicity of their associated habit hierarchies (low D and high D words), were assumed to interact differently with anxiety level. The results were as follows:

1. HA subjects tended to give longer reaction times in the Word Association test, independent of the D value of the word.

2. No relationship between anxiety level and number of reproduction faults or number of rare responses was found, contradicting earlier findings by Sarason (1959).

3. Over all 99 words, constituting the total Word Association test, it was found that HA subjects had longer reaction times than LA subjects.

The results were interpreted as supporting a habit interference view of anxiety, although it was suggested that a dichotomous view of the drive vs. habit issue is not necessary. Sarason's (1959) earlier findings were interpreted as being due more to response censorship than to genuine differences in verbal behavior.

REFERENCES

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.

HEINEMAN, C. E. A forced-choice form of the Taylor anxiety scale. *J. consult. Psychol.*, 1953, **17**, 447–455.

LAFFAL, J. The learning and retention of words with association disturbances. *J. abnorm. soc. Psychol.*, 1952, **47**, 454–462.

LAFFAL, J. Response faults in word association as a function of response entropy. *J. abnorm soc. Psychol.*, 1955, **50**, 265–270.

ROSEN, E., & RUSSELL, W. A. Frequency characteristics of successive word associations. *Amer. J. Psychol.*, 1957, **70**, 120–122.

RUSSELL, W. A., & JENKINS, J. J. The complete Minnesota norms for responses to 100 words from the Kent-Rosanoff Word Association Test. *USN Off. Naval Res. tech. Rep.*, 1954, No. 11. (Contract No. N8)

SARASON, I. G. Relationships of measures of anxiety and experimental instructions to word association performance. *J. abnorm. soc. Psychol.*, 1959, **59**, 37–42.

SPENCE, K. W. A theory of emotionally based drive (D) and its relation to performance in simple learning situations. *Amer. Psychologist*, 1958, **13**, 131–141.

SPENCE, K. W., FARBER, I. E., & MCFANN, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associates learning. *J. exp. Psychol.*, 1956, **52**, 296–305.

TAYLOR, JANET. The effects of anxiety level and psychological stress on verbal learning. *J. abnorm. soc. Psychol.*, 1958, **57**, 55–60.

(Received March 10, 1960)

Journal of Abnormal and Social Psychology
1961, Vol. 62, No. 2, 472-474

RELATIONSHIPS BETWEEN SELF-ACCEPTANCE AND SOCIOMETRIC CHOIC

HAYNE W. REESE
University of Buffalo

SEVERAL studies (e.g., Fey, 1955; McIntyre, 1952) found no significant relationship between self-acceptance and acceptance by others, but Miyamoto and Dornbusch (1956) found higher acceptance by others for subjects with high self-concepts than those with low self-concepts, and Marshall (1958) found in an experimental study that a negative shift in self-acceptance followed unfavorable reactions by the peer group, though no change followed favorable reactions. The conflicting nature of the evidence may result in part from the use of different definitions of self-acceptance. In the present study two measures of self-acceptance were used, and their relationships with several sociometric measures were investigated.

METHOD

Subjects. The subjects were 507 children from the fourth, sixth, and eighth grades of the Clarence, New York, school system.[1] The data of 66 of these subjects were eliminated, 30 subjects missed one or more of the tests, 9 subjects obtained negative discrepancy scores, and 27 subjects named no best friend in their class. Additional subjects, as indicated below, were eliminated in a random manner to obtain proportional cell frequencies for the analyses of variance.

Tests and procedure. The tests were Lipsitt's (1958) Self-Concept scale and his Ideal-Self scale, a "Rate" Sociometric scale similar to one described by McCandless, Castaneda, and Palermo (1956), and a "Check-List" Sociometric scale. Lipsitt's Self-Concept scale contains 22 trait-descriptive adjectives, each prefaced by the phrase "I am . . ." and followed by a five-point rating scale with the titles "not at all," "not very often," "some of the time," "most of the time," and "all of the time." The Ideal-Self scale contains the same adjectives, prefaced by the phrase "I would like to be . . . ," and the same rating categories.

The Rate Sociometric scale contained the names of every same-sexed member of a given class, with a five-point scale beside each name. The subjects were instructed to check 1 if the child named was a "best friend"; 2 if he was a friend, but not a best friend; 3 if the subject did not know the child very well or neither liked nor disliked him; 4 if the subject disliked him more than he liked him; and 5 if the subject did not like the child.

The Check-List Sociometric scale contained th names of every same-sexed member of a given clas and the subjects were instructed to check the names their "best friends." Only the children who were bot checked and rated 1 were counted as the subject's bes friends.

The tests were given in two sessions by the teach in their regular classes. If a child missed a session, was tested individually if he returned to school within week; otherwise his data were eliminated. If t discrepancy score, i.e., the ideal-self minus the self-concept score, was negative, the subject's data wer eliminated (Lipsitt, 1958). The data of the subjects wh had no "best friends," as defined above, were also elim nated.

For the analyses of the data, the same-sexed subject within each grade were divided into three groups c the basis of self-concept scores, and also on the ba of discrepancy scores. Group L consisted of the lowe third, Group M the middle third, and Group H th highest third. It was assumed that the differential responses to Lipsitt's (1958) scales reflected differen degrees of self-acceptance, though they may reflec differential tendencies to respond in socially favorab terms. The empirical findings are not affected by possible social desirability component of Lipsitt's sc though theoretical interpretation might be affected

RESULTS

Groups based on self-concept scores. To obt proportional cell frequencies the data of 33 s jects were eliminated, leaving 408 subjects in groups based on self-concept scores. There 17 boys and 17 girls in each group in the fo and sixth grades, and 34 of each sex in each g in the eighth grade. Girls obtained a significa higher mean self-concept score than boys (see column of Table 2); the sex difference was lar in Group L than in Groups M and H (t = 2.4 4.47),[2] and larger in Group M than in Group (t = 2.02). The sex difference was smaller in Gr 4 than in Grades 6 and 8 (t = 2.15, 2.61), but latter two grades did not differ significantly in size of the sex difference (t < 1.00).

The data provided two measures of acceptan of others, the mean of the ratings the subject ga others, and the percentage he named as friends. Since the product-moment correlation tween these measures was −.65 (indicating a

[1] The writer is indebted to John Rosenbach, Director of Pupil Personnel, Clarence Central Schools; Delbert H. Repp, Principal, Clarence Central Jr. High School; John O'Neil, Principal, Sheridan Hill Elementary School; Ben Macey, Principal, Parker Elementary School; and their office and teaching staffs, for their fine cooperation and assistance in this study.

[2] The error term of each t was based on the wit cells MS of the analysis of variance, and the df of the same as the df of the within-cells MS. Unless ot wise indicated, each t reported is significant at the level or better.

ive relationship), and they were related to self-acceptance in essentially the same way, only the findings for acceptance of others measured by the mean rating are reported. The first line of Table 1 gives these means. The curvilinear trend was significant, as indicated in the second column of Table 2, and tests of the differences between individual means showed that Groups L and M differed significantly ($t = 2.65$), but Group H was not significantly different from L or M ($t = 1.45$, 1.20). Analysis of the simple effects in the Sex × Grade interaction showed that the sixth and eighth grade girls obtained lower means than the boys in these grades ($t = 3.56$, 2.34), but the sex difference in the fourth grade was in the opposite direction and not reliable ($t = 1.53$).

The second line of Table 1 gives the mean acceptance by same-sexed classmates, or "acceptance by others." The curvilinear trend, paralleling that for acceptance of others, was significant, and t tests showed that each group was significantly different from the other groups ($t = 4.44$ for L vs. M; 2.14 for L vs. H; 2.30 for M vs. H). The girls were more highly accepted by others than boys were in the sixth and eighth grades ($t = 3.20$, 2.18), but the sex difference in the fourth grade was in the opposite direction and not significant ($t < 1.00$).

The product-moment correlation between acceptance by others and acceptance by best friends was .63, and the latter measure was related to self-acceptance in essentially the same way as the former, though less strongly (see Tables 1 and 2). Groups L and M differed significantly in acceptance by best friends ($t = 3.46$), but Group H was not significantly different from Group L or M ($t = 1.77$, 1.69), and none of the simple effects in the Sex × Grade interaction was significant.

Acceptance by others was positively related to the two measures of acceptance of others; the product-moment correlation with the mean rating given to others was .20 ($p < .001$), and the correlation with the percentage measure was −.20 (indicating a positive relationship). Acceptance by best friends was not significantly related to the mean rating given to others ($r = -.01$), but was negatively related to the percentage measure ($r = .10$, $p < .05$).

Groups based on discrepancy scores. To obtain proportional cell frequencies in the groups formed on the basis of discrepancy scores, the data of 45 subjects were eliminated, leaving 16 boys and 16 girls in each fourth and sixth grade group, and 34 of each sex in each eighth grade group. The sixth grade boys obtained a higher mean discrepancy score than sixth grade girls ($t = 2.55$), but the sex differences in the fourth and eighth grades were not significant (each $t < 1.00$). Group L girls obtained a higher mean discrepancy than Group L boys ($t = 2.13$), but the trend was reversed in Group H ($t = 4.66$), and the means were equal in Group M.

TABLE 1

MEAN SOCIOMETRIC SCORES OF GROUPS BASED ON SELF-CONCEPT SCORES

Sociometric Measure	Group		
	L	M	H
Acceptance of Others	2.23	2.10	2.16
Acceptance by Others	2.30	2.02	2.17
Acceptance by Best Friends	1.75	1.47	1.60

Note.—Sex and grade have been collapsed since neither interacted significantly with groups. A low mean indicates high acceptance.

TABLE 2

SUMMARY OF ANALYSES OF VARIANCE OF GROUPS BASED ON SELF-CONCEPT SCORES

Source	*df*	Self-Concept Scores		Acceptance of Others		Acceptance by Others		Acceptance by Friends	
		MS	*F*	*MS*	*F*	*MS*	*F*	*MS*	*F*
Level (L)	2	11982	704.82****	.63	3.50*	2.73	9.75****	2.62	5.95***
Sex (S)	1	1491	87.71****	1.28	7.11**	2.49	8.89***	1.05	2.39
Grade (G)	2	3	<1.00	.04	<1.00	.04	<1.00	2.18	4.95**
L × S	2	168	9.88****	.17	<1.00	.08	<1.00	.01	<1.00
L × G	4	31	1.82	.05	<1.00	.44	1.57	.94	2.14
S × G	2	62	3.65*	1.20	6.67***	.84	3.00*	1.50	3.41*
L × S × G	4	34	2.00	.40	2.22	.58	2.07	.60	1.36
Within	390	17		.18		.28		.44	

Note.—"Level" refers to groups formed on the basis of self-concept scores. The last column summarizes the analysis of acceptance by best friends.

* $p < .05$.

** $p < .01$.

*** $p < .005$.

**** $p < .001$.

Analyses of the sociometric measures indicated that there was no significant Group effect; there were significant sex and grade differences, but these were not consistent between different measures.

SUMMARY AND CONCLUSIONS

Lipsitt's (1958) Self-Concept and Ideal-Self scales and two Sociometric scales were given to fourth, sixth, and eighth grade subjects.

1. Acceptance of others, acceptance by others, and acceptance by best friends were curvilinearly related to self-concept scores, with highest acceptance in a group with moderate self-concept scores and lowest in a group with low self-concept scores.

2. The trends were not significantly different in different grades or sexes.

3. Acceptance by others was more strongly related to self-concept than was acceptance by best friends.

4. The sociometric measures were not significantly related to the discrepancy between ideal-self and self-concept scores, suggesting that the discrepancy score obtained by subtraction may not be a valid measure of self-acceptance.

REFERENCES

FEY, W. F. Acceptance by others and its relation to acceptance of self and others: A revaluation. *J abnorm. soc. Psychol.*, 1955, **50**, 274–276.

LIPSITT, L. P. A self-concept scale for children and its relationship to the children's form of the manifest anxiety scale. *Child Develpm.*, 1958, **29**, 463–472.

MCCANDLESS, B. R., CASTANEDA, A., & PALERMO, D. S. Anxiety in children and social status. *Child Develpm.*, 1956, **27**, 385–391.

MCINTYRE, C. J. Acceptance by others and its relation to acceptance of self and others. *J. abnorm. soc. Psychol.*, 1952, **47**, 624–625.

MARSHALL, R. J. Variation in self-attitudes and attitudes toward others as a function of peer group appraisals. Unpublished doctoral dissertation, University of Buffalo, 1958.

MIYAMOTO, S. F., & DORNBUSCH, S. M. A test of interactionist hypotheses of self-conception. *Amer. J. Sociol.*, 1956, **61**, 399–403.

(Received March 27, 1960)